PACKING ELLIPSOIDS WITH OVERLAP*

CAROLINE UHLER† AND STEPHEN J. WRIGHT ‡

Abstract. The problem of packing ellipsoids of different sizes and shapes into an ellipsoidal container so as to minimize a measure of overlap between ellipsoids is considered. A bilevel optimization formulation is given, together with an algorithm for the general case and a simpler algorithm for the special case in which all ellipsoids are in fact spheres. Convergence results are proved and computational experience is described and illustrated. The motivating application — chromosome organization in the human cell nucleus — is discussed briefly, and some illustrative results are presented.

Key words. ellipsoid packing, trust-region algorithm, semidefinite programming, chromosome territories.

AMS subject classifications. 90C22, 65K10, 92C37

1. Introduction.

Shape packing problems have been a popular area of study in discrete mathematics over many years. Typically, such problems pose the question of how many uniform objects can be packed without overlap into a larger container, or into a space of infinite extent with maximum density. In ellipsoid packing problems, the smaller shapes are taken to be ellipsoids of known size and shape. In three dimensions, the ellipsoid packing problem has become well known in recent years, due in part to colorful experiments involving the packing of M&Ms [8].

Finding densest ellipsoid packings is a difficult computational problem. Most studies concentrate on the special case of sphere packings, with spheres of identical size. Here, optimal densities have been found for the infinite Euclidean space of dimensions two and three. In two dimensions, the densest circle packing is given by the hexagonal lattice (see [22]), where each circle has six neighbors. The density of this packing (that is, the proportion of the space filled by the circles) is $\pi/\sqrt{12}$. In dimension three, it has been proven recently by Hales [10] that the face-centered cubic (FCC) lattice achieves the densest packing. In this arrangement, every sphere has 12 neighboring spheres and the density is $\pi/\sqrt{18}$. For dimensions higher than 3, the problem of finding the densest sphere packing is still open.

A problem related to sphere packing is *sphere covering*. Here, the goal is to find an arrangement that covers the space with a set of uniform spheres, as economically as possible. Overlap is not only allowed in these arrangements, but inevitable. The density is defined similarly to sphere packing (that is, the total volume of the spheres divided by the volume covered), but now we are interested in finding an arrangement of *minimal* density. In two dimensions, as for circle packing, the optimal circle covering is given by the regular hexagonal arrangement. However, the thinnest sphere covering in dimension 3 is given not by the FCC lattice, but by the body-centered cubic (BCC) lattice. In this arrangement, every sphere intersects with fourteen neighboring spheres; see for example [20].

In this paper we study a problem that falls between ellipsoid packing and covering. Given a set of ellipsoids of diverse size and shape, and a larger enclosing ellipsoid to

*Version of February 7, 2013. Research supported by NSF Grants DMS-0914524, DMS-0906818, and DMS-1216318, and DOE Grant DE-SC0002283.

†IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria. caroline.uhler@ist.ac.at

‡Computer Sciences Department, 1210 W. Dayton Street, University of Wisconsin, Madison, WI 53706, USA. swright@cs.wisc.edu

which the smaller ellipsoids are all restricted, we seek an arrangement that minimizes some measure of total overlap between ellipsoid pairs.

Our formulation is motivated by chromosome organization in human cell nuclei. In biological sciences, the study of chromosome arrangements and their functional implications is an area of great current interest. The territory occupied by each chromosome can be modeled as an ellipsoid, different chromosomes giving rise to ellipsoids of different size. The enclosing ellipsoid represents a cell nucleus, the size and shape of which differs across cell types. Overlap between chromosome territories has biological significance: It allows for interaction and co-regulation of different genes. Also of key significance are the DNA-free interchromatin channels that allow access by regulatory factors to chromosomes deep inside a cell nucleus. Smaller nuclei tend to have tighter packings, so that fewer channels are available, and the chromosomes packed closest to the center may not be accessible to regulatory factors.

The arrangement of chromosome territories is neither completely random nor deterministic. Certain features of the arrangement are believed to be conserved during evolution [21], but can change during such processes as cell differentiation and cancer development [14]. In general, smaller and more gene-dense chromosomes are believed to be found closer to the center of the nucleus [1], and heterologous chromosomes tend to be nearer to each other than homologous pairs [11]. For further background on chromosome arrangement properties, see [6, 24].

A major goal of this paper is to determine whether the experimental observations made to date about chromosome organization can be explained in terms of simple geometrical principles, such as minimal overlap. The minimum-overlap principle appears to be consistent with the tendency of chromosome territories to exploit the whole volume of the nucleus, to make the DNA-free channels as extensive as possible. Our formulation also includes features to discourage close proximity of homologous pairs.

The remainder of the paper is organized as follows. In Section 2, we outline the mathematical formulation, define notation, and state a key technical result concerning algebraic formulations of ellipsoidal containment. In Section 3, we study the special case of finding a minimal overlap configuration of *spheres* inside a convex container. We describe a simple iterative procedure based on convex linearized approximations that produces convergence to stationary points of the minimal-overlap problem. We show through simulations that our algorithm can be used to recover known optimal circle and sphere packings. In Section 4, we generalize our optimization procedure to ellipsoid packing, introducing trust-region stabilization and proving convergence results. Section 5 describes the application of our algorithms to chromosome arrangement.

Notation. When A and B are two symmetric matrices, the relation $A \preceq B$ indicates that $B-A$ is positive semidefinite, while $A \prec B$ denotes positive definiteness of $B - A$. Similar definitions apply for $\succeq$ and $\succ$.

Let $\mathcal{X}$ be a finite-dimensional vector space over the reals $\mathbb{R}$ endowed with inner product $\langle \cdot, \cdot \rangle$. (The usual Euclidean space $\mathbb{R}^n$ with inner product $\langle x, y \rangle := x^T y$ and the space of symmetric matrices $\mathcal{S}\mathbb{R}^{n\times n}$ with inner product $\langle X, Y \rangle := \mathrm{trace}(XY)$ are two examples of particular interest in this paper.) Given two elements x and y of $\mathcal{X}$, we write $x \perp y$ if $\langle x, y \rangle = 0$.

Given a closed convex subset $\Omega \subset \mathcal{X}$, the normal cone to Ω at a point $x \in \Omega$ is defined as

$$N_\Omega(x) := \{v \in \mathcal{X} \mid \langle v, y - x \rangle \leq 0 \text{ for all } y \in \Omega\}. \tag{1.1}$$

(We follow the convention that $N_\Omega(x) = \emptyset$ if $x \notin \Omega$.) We use ∂f to denote the Clarke subdifferential of the function $f : \mathcal{X} \to \mathbb{R}$. In defining this quantity, we follow Borwein and Lewis [2, p. 124] by assuming Lipschitz continuity of f at x, and defining the Clarke directional derivative as follows:

$$f^\circ(x; h) := \limsup_{y \to x, t \downarrow 0} \frac{f(y + th) - f(y)}{t}.$$

The Clarke subdifferential is then

$$\partial f(x) := \{v \in \mathcal{X} \mid \langle v, h\rangle \leq f^\circ(x; h), \text{ for all } h \in \mathcal{X}\}. \tag{1.2}$$

When f is convex (in addition to Lipschitz continuous), this definition coincides with the usual subdifferential from convex analysis, which is

$$\partial f(x) := \{v \in \mathcal{X} \mid f(y) \geq f(x) + \langle v, y - x\rangle \text{ for all } y \in \text{dom } f\}$$

(see [5, Proposition 2.2.7]).

2. Problem Description and Preliminaries. An ellipsoid $\mathcal{E} \subset \mathbb{R}^n$ can be specified in terms of its center $c \in \mathbb{R}^n$ and a symmetric positive definite eccentricity matrix $S \in S\mathbb{R}^{n\times n}$. We can write

$$\mathcal{E} := \{x \in \mathbb{R}^n \mid (x - c)^T S^{-2}(x - c) \leq 1\} = \{c + Su \mid \|u\|_2 \leq 1\}. \tag{2.1}$$

It is often convenient to work with the quantity $\Sigma := S^2$ (also symmetric positive definite), and thus to rewrite the definition (2.1) as

$$\mathcal{E} := \{x \in \mathbb{R}^n \mid (x - c)^T \Sigma^{-1}(x - c) \leq 1\}. \tag{2.2}$$

In most of the paper, we assume that $n = 3$, that is, the ellipsoids are three-dimensional. This suffices for our application, though generalizations to arbitrary dimensions are fairly straightforward. (In the special case of spheres (Section 3), no modifications to the algorithm or analysis are needed for general n.)

The eigenvalues of S are the lengths of the principal semi-axes of $\mathcal{E}$; we denote these by r_1, r_2, and r_3, and assume that these three positive quantities are arranged in nonincreasing order. It follows that the eigenvalues of Σ are r_1^2, r_2^2, and r_3^2, and that the matrices S and Σ have the form

$$S = Q \begin{bmatrix} r_1 & 0 & 0 \\ 0 & r_2 & 0 \\ 0 & 0 & r_3 \end{bmatrix} Q^T, \qquad \Sigma = Q \begin{bmatrix} r_1^2 & 0 & 0 \\ 0 & r_2^2 & 0 \\ 0 & 0 & r_3^2 \end{bmatrix} Q^T,$$

for some orthogonal matrix Q, which determines the orientation of the ellipsoid.

In this paper, we are given the semi-axis lengths r_{i1}, r_{i2}, and r_{i3} for a collection of N ellipsoids $\mathcal{E}_i$, $i = 1, 2, \ldots, N$. The goal is to specify centers c_i and matrices S_i for these ellipsoids, such that

(a) $\mathcal{E}_i \subset \mathcal{E}$, for some fixed ellipsoidal container $\mathcal{E}$;
(b) The eigenvalues of S_i are r_{i1}, r_{i2}, and r_{i3}, for $i = 1, 2, \ldots, N$;
(c) Some measure of volumes of the pairwise overlaps $\mathcal{E}_i \cap \mathcal{E}_j$, $i, j = 1, 2, \ldots, N$, $i \neq j$, is minimized.

In the following subsections, we give more specific formulations of (c), first for the case in which all $\mathcal{E}_i$ are spheres (that is, $r_{i1} = r_{i2} = r_{i3}$, $i = 1, 2, \ldots, N$) and then for the general case. For now, we note that a crucial element in formulating these problems is ellipsoidal containment, that is, algebraic conditions that ensure that one given ellipsoid is contained in another. This is the subject of the following lemma, which is a simple application of the S-procedure (see [4, Appendix B.2], [18], and [3]).

LEMMA 2.1. *Define two ellipsoids as follows:*

$$\mathcal{E} = \{x \in \mathbb{R}^3 \mid (x-c)^T S^{-2}(x-c) \leq 1\} = \{c + Su \mid \|u\|_2 \leq 1\},$$
$$\bar{\mathcal{E}} = \{x \in \mathbb{R}^3 \mid (x-\bar{c})^T \bar{S}^{-2}(x-\bar{c}) \leq 1\} = \{\bar{c} + \bar{S}u \mid \|u\|_2 \leq 1\}.$$

The containment condition $\bar{\mathcal{E}} \subset \mathcal{E}$ can be represented as the following linear matrix inequality (LMI) in parameters $\bar{c}$, $\bar{S}$, c, and S^2: There exists $\lambda \in \mathbb{R}$ such that

$$\begin{pmatrix} -\lambda I & 0 & \bar{S} \\ 0 & \lambda - 1 & (\bar{c}-c)^T \\ \bar{S} & \bar{c}-c & -S^2 \end{pmatrix} \preceq 0. \tag{2.3}$$

Proof. The condition $\bar{\mathcal{E}} \subset \mathcal{E}$ can be expressed as

$$(\bar{c} + \bar{S}u - c)^T S^{-2}(\bar{c} + \bar{S}u - c) \leq 1 \quad \text{for all } u \text{ such that } \|u\|_2 \leq 1.$$

By multiplying out this inequality we get

$$u^T \bar{S} S^{-2} \bar{S} u + 2u^T \bar{S} S^{-2}(\bar{c}-c) + (\bar{c}-c)^T S^{-2}(\bar{c}-c) - 1 \leq 0$$

for all u such that $u^T u - 1 \leq 0$. By applying the S-procedure, we find that this is equivalent to the existence of $\lambda > 0$ such that

$$\begin{pmatrix} \bar{S} S^{-2} \bar{S} & \bar{S} S^{-2}(\bar{c}-c) \\ (\bar{c}-c)^T S^{-2} \bar{S} & (\bar{c}-c)^T S^{-2}(\bar{c}-c) - 1 \end{pmatrix} \preceq \lambda \begin{pmatrix} I & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.4}$$

This expression is not linear in the variables $\bar{c}$, $\bar{S}$, c, and S^2, but an elementary Schur complement argument shows equivalence to the linear matrix inequality (2.3). ∎

As one special case, the condition $\mathcal{E}_i \subset \mathcal{E}$, where $\mathcal{E}_i$ is a sphere with center c_i and radius r_i and $\mathcal{E}$ is an ellipsoid centered at 0 with matrix S, can be represented as the LMI:

$$\begin{pmatrix} -\lambda_i I & 0 & r_i I \\ 0 & \lambda_i - 1 & c_i^T \\ r_i I & c_i & -S^2 \end{pmatrix} \preceq 0. \tag{2.5}$$

3. Sphere Packing. We give a problem formulation for the case in which all enclosed shapes are spheres (of arbitrary dimension), and present a successive approximation algorithm that is shown to accumulate or converge to a stationary point of the formulation. Some examples of results obtained with this approach are described at the end of the section.

3.1. Formulation and Algorithm. When the inscribed objects are spheres, the variables in the problem are the centers $c_i \in \mathbb{R}^n$, $i = 1, 2, \ldots, N$, which we aggregate as follows:

$$c := (c_1, c_2, \ldots, c_N). \tag{3.1}$$

The radii r_i, $i = 1, 2, \ldots, N$ are given. We express the containment condition for each sphere as follows:

$$\mathcal{E}_i \subset \mathcal{E} \quad \Leftrightarrow \quad c_i \in \Omega_i, \tag{3.2}$$

where Ω_i is a closed, bounded, convex set with nonempty interior. When $\mathcal{E}$ is a sphere of radius R centered at 0, we have $\Omega_i := \{c_i \,:\, \|c_i\| \leq R - r_i\}$. Otherwise, we can define Ω_i implicitly by Lemma 2.1; see in particular (2.5).

A simple measure for the overlap between two spheres $\mathcal{E}_i$ and $\mathcal{E}_j$ is the diameter of the largest sphere inscribed into the intersection, which we denote by an auxiliary variable ξ_{ij}:

$$\xi_{ij} := \max(0, (r_i + r_j) - \|c_i - c_j\|_2), \qquad \xi := (\xi_{ij})_{1 \leq i < j \leq N}. \tag{3.3}$$

Our minimum-overlap problem can thus be formulated as follows:

$$\begin{aligned}
\min_{c,\xi} \quad & H(\xi) && && \text{(3.4a)}\\
\text{subject to} \quad & (r_i + r_j) - \|c_i - c_j\|_2 \leq \xi_{ij} && \text{for } 1 \leq i < j \leq N && \text{(3.4b)}\\
& 0 \leq \xi, && && \text{(3.4c)}\\
& c_i \in \Omega_i, && \text{for } i = 1, \ldots, N, && \text{(3.4d)}
\end{aligned}$$

where (3.4c) denotes the entrywise condition $\xi_{ij} \geq 0$, $1 \leq i < j \leq N$. The objective $H : \mathbb{R}_+^{n(n-1)/2} \to \mathbb{R}_+$ is chosen to satisfy the following assumption.

ASSUMPTION 1. *The function $H : \mathbb{R}_+^{n(n-1)/2} \to \mathbb{R}_+$ is convex and continuous, with the following additional properties:*

(a) $H(0) = 0$;

(b) $H(\xi) > 0$ whenever $\xi \neq 0$;

(c) $0 \leq \bar{\xi} \leq \xi \Rightarrow H(\bar{\xi}) \leq H(\xi)$.

Assumption 1 is satisfied, for example, by the norms $H(\xi) = \|\xi\|_1$, $H(\xi) = \|\xi\|_2$, and $H(\xi) = \|\xi\|_\infty = \max_{1 \leq i < j \leq N} |\xi_{ij}|$. In the application to be discussed below, we prefer the overlaps in the overlapping ellipsoids to be roughly the same size; for this purpose, the ℓ_2 and ℓ_∞ norms are the most appropriate.

Although the objective (3.4a) and containment constraints (3.4d) are convex, the problem (3.4) is nonconvex, due to the constraints (3.4b). A point is *Clarke-stationary* for (3.4) if the following conditions are satisfied, for some $\lambda_{ij} \in \mathbb{R}$, $1 \leq i < j \leq N$:

$$\begin{aligned}
0 \leq g_{ij} - \lambda_{ij} \perp \xi_{ij} \geq 0 \ \text{ for some } g_{ij} \in \partial_{\xi_{ij}} H(\xi), \qquad & 1 \leq i < j \leq N, && \text{(3.5a)}\\
\sum_{j=i+1}^{N} \lambda_{ij} w_{ij} - \sum_{j=1}^{i-1} \lambda_{ji} w_{ji} \in N_{\Omega_i}(c_i), \qquad & i = 1, 2, \ldots, N, && \text{(3.5b)}\\
0 \leq \xi_{ij} + \|c_i - c_j\| - (r_i + r_j) \perp \lambda_{ij} \geq 0, \qquad & 1 \leq i < j \leq N, && \text{(3.5c)}\\
\text{where } \|w_{ij}\|_2 \leq 1, \quad \text{with } \ w_{ij} = \frac{c_i - c_j}{\|c_i - c_j\|_2} \text{ when } c_i \neq c_j, \quad & 1 \leq i < j \leq N. && \text{(3.5d)}
\end{aligned}$$

Condition (3.5d) defines w_{ij} to be in the subdifferential of $\|c_i - c_j\|_2$ with respect to c_i. See (1.1) for the definition of the normal cone in (3.5b).

We now develop an algorithm that seeks a local solution of (3.4), by formulating a sequence of convex approximations in which the key feature is linearization of the nonconvex constraint (3.4b) around the current iterate. Because of the special

properties of this problem, we need not apply the usual safeguards for this successive approximation approach, such as trust regions or line searches. Decrease of the objective at each iteration and accumulation of the iteration sequence at first-order points of the problem (3.4) can be proved in the absence of these features, as we show below. However, for purposes of stabilizing the iterates generated by the method, it may be desirable to place a uniform bound on the length of each step, and we do so in our implementations.

The linearization of (3.4) around the current iterate c^- is defined as follows:

$$P(c^-) := \min_{c,\bar{\xi}} H(\bar{\xi}) \tag{3.6a}$$

$$\text{subject to} \quad (r_i + r_j) - z_{ij}^T(c_i - c_j) \le \bar{\xi}_{ij}, \qquad \text{for } 1 \le i < j \le N, \tag{3.6b}$$

$$0 \le \bar{\xi}, \tag{3.6c}$$

$$c_i \in \Omega_i, \qquad \text{for } i = 1, \ldots, N, \tag{3.6d}$$

$$\text{where} \quad z_{ij} := \begin{cases} (c_i^- - c_j^-)^T / \|c_i^- - c_j^-\| & \text{when } c_i^- \neq c_j^- \\ 0 & \text{otherwise.} \end{cases} \tag{3.6e}$$

This problem is convex, with affine constraints except for the inclusion (3.6d), which can be satisfied strictly when each Ω_i is closed, bounded, and convex, with nonempty interior. Hence (see for example [19, Theorem 28.2, Corollary 28.3.1]), its solutions are characterized by the following KKT conditions: There exist λ_{ij}, $1 \le i < j \le N$ such that

$$0 \le g_{ij} - \lambda_{ij} \perp \bar{\xi}_{ij} \ge 0 \ \text{ for some } g_{ij} \in \partial_{\bar{\xi}_{ij}} H(\bar{\xi}), \qquad 1 \le i < j \le N, \tag{3.7a}$$

$$\sum_{j=i+1}^{N} \lambda_{ij} z_{ij} - \sum_{j=1}^{i-1} \lambda_{ji} z_{ji} \in N_{\Omega_i}(c_i), \qquad i = 1, 2, \ldots, N, \tag{3.7b}$$

$$0 \le \bar{\xi}_{ij} + z_{ij}^T(c_i - c_j) - (r_i + r_j) \perp \lambda_{ij} \ge 0, \qquad 1 \le i < j \le N. \tag{3.7c}$$

We can use a compactness argument to verify that solutions to (3.6) are attained. The vector of feasible centers c is restricted to a compact set, by the assumed properties of $\Omega_1, \Omega_2, \ldots, \Omega_N$. By using (3.6b) we can define effective upper bounds on the variables $\bar{\xi}_{ij}$ as follows:

$$\bar{\xi}'_{ij} := \max\left(0, \sup_{c_i \in \Omega_i,\, c_j \in \Omega_j} (r_i + r_j) - z_{ij}^T(c_i - c_j)\right).$$

(For any feasible c, and given any $\bar{\xi}$ satisfying (3.6b), we can always replace $\bar{\xi}$ by an alternative feasible point $\bar{\xi}'' \in [0, \bar{\xi}']$ without increasing the value of H, by property (c) of Assumption 1.) Thus, the problem (3.6) reduces to minimization of a continuous convex function over a compact set, for which existence of a solution is guaranteed.

Algorithm 1 is our algorithm, based on the subproblem (3.6). To analyze convergence properties of this method, we start with basic results about stationary points and about the changes in H at each iteration of Algorithm 1.

LEMMA 3.1. *Suppose that the sets Ω_i in (3.4) are closed, bounded, and convex, with a nonempty interior, and that Assumption 1 holds. Then the following claims are true.*

(i) If the point (c^k, ξ^k) satisfies the optimality conditions (3.7) for the subproblem $P(c^k)$ defined by (3.6), then (c^k, ξ^k) satisfies the stationarity conditions (3.5) for the problem (3.4).

Algorithm 1 Packing Spheres by Minimizing Overlap

Given r_i, $i = 1, 2, \ldots, N$ and Ω_i closed, convex, bounded with nonempty interior;
Choose $c^0 \in \Omega_1 \times \Omega_2 \times \cdots \times \Omega_N$;
for $k = 0, 1, 2, \ldots$ **do**
  Solve $P(c^k)$ defined by (3.6) to obtain $(c^{k+1}, \bar{\xi}^{k+1})$;
  if $H(\bar{\xi}^{k+1}) = H(\xi^k)$ **then**
    stop and return c^k;
  end if
  Set $\xi_{ij}^{k+1} = \max(0, (r_i + r_j) - \|c_i^{k+1} - c_j^{k+1}\|)$ for $1 \le i < j \le N$;
end for

(ii) If the point (c^k, ξ^k) satisfies the stationarity conditions (3.5) for the problem (3.4) and in addition $c_i^k \neq c_j^k$ for all $1 \le i < j \le N$, then (c^k, ξ^k) satisfies the optimality conditions (3.7) for the subproblem $P(c^k)$ defined by (3.6).

(iii) If (c^k, ξ^k) does not satisfy the stationarity conditions (3.5), then $H(\bar{\xi}^{k+1}) < H(\xi^k)$.

(iv) For each k we have $H(\xi^k) \le H(\bar{\xi}^k)$.

Proof.

(i) If (c^k, ξ^k) satisfies the optimality conditions (3.7) for $P(c^k)$, then by setting $w_{ij} = z_{ij}$ in (3.5), we see that these conditions are also satisfied with the same values of g_{ij} and λ_{ij}. (We have made the particular choice $w_{ij} = 0$ when $c_i^k = c_j^k$.)

(ii) If the conditions (3.5) are satisfied at (c^k, ξ^k) with $c_i^k \neq c_j^k$ for all i, j with $1 \le i < j \le N$, then $w_{ij} = (c_i^k - c_j^k)/\|c_i^k - c_j^k\|$ for all such i, j. Thus by noting that $z_{ij} = w_{ij}$ for all such i, j, we can verify using the same values of g_{ij} and λ_{ij} that (c^k, ξ^k) satisfies the optimality conditions (3.7), and therefore is a solution of $P(c^k)$.

(iii) Note that the point (c^k, ξ^k) is feasible for the subproblem (3.6) (with $c^- = c^k$), so its optimal objective satisfies $H(\bar{\xi}^{k+1}) \le H(\xi^k)$. Since (c^k, ξ^k) does not satisfy the stationarity conditions (3.5), however, part (i) implies that it cannot be a *solution* of (3.6), which implies that in fact $H(\bar{\xi}^{k+1}) < H(\xi^k)$, as claimed.

(iv) By using the fact that $\|z_{ij}^{k-1}\|_2 \le 1$, we have for all $k \ge 1$ and all i, j with $1 \le i < j \le N$ that

$$\xi_{ij}^k = \max(r_i + r_j - \|c_i^k - c_j^k\|_2, 0) \le \max(r_i + r_j - (z_{ij}^{k-1})^T (c_i^k - c_j^k), 0) \le \bar{\xi}_{ij}^k.$$

The result now follows immediately from Assumption 1(c).

□

Note that in the case of coinciding centers, i.e. $c_i = c_j$ for some $i \neq j$, the stationarity conditions for (3.4) and (3.6) are not equivalent. This observation yields the intriguing property — unusual in algorithms based on linear approximations — that Algorithm 1 may be able to move away from a stationarity point for (3.4). That is, if (c^k, ξ^k) satisfies (3.5) but there is some pair (i, j) with $i \neq j$ and $c_i^k = c_j^k$, then by setting $z_{ij} = 0$, the subproblem (3.7) may yield a solution $(c^{k+1}, \bar{\xi}^{k+1})$ with $H(\bar{\xi}^{k+1}) < H(\xi^k)$, and thus (by Lemma 3.1 (iv)) the next iterate will satisfy $H(\xi^{k+1}) < H(\xi^k)$. Note too that the proof of Lemma 3.1 (iv) still holds if z_{ij}^k is chosen to be *any* vector with $\|z_{ij}^k\|_2 \le 1$ when $c_i^k = c_j^k$. Hence, random choices for z_{ij}^k in this situation could be used in place of our choice $z_{ij}^k = 0$ above, leading to some interesting algorithmic

possibilities for avoiding coincident centers and moving away from stationary points. Since coincident centers rarely arise in the cases of interest, however, we do not pursue these possibilities.

We now prove the main convergence result for Algorithm 1.

THEOREM 3.2. *Suppose that the sets Ω_i in (3.4) are closed, bounded, and convex, with a nonempty interior, and that Assumption 1 holds. Then Algorithm 1 either terminates at a stationary point for (3.4), or else generates an infinite sequence $\{c^k\}$ for which all accumulation points $\hat{c}$ are either stationary points for (3.4), or else have $\hat{c}_i = \hat{c}_j$ for some pair (i, j) with $1 \leq i < j \leq N$.*

Proof. Lemma 3.1 (iii) says that termination can occur only if (c^k, ξ^k) satisfies the stationarity conditions (3.5). Hence, we need to consider only the case of an infinite sequence of iterates $\{c^k\}$. Suppose for contradiction that there is an accumulation point $\hat{c}$ for this sequence such that $\hat{c}_i \neq \hat{c}_j$ for all (i, j) but $\hat{c}$ is not stationary for (3.4). Considering the problem $P(\hat{c})$ defined by (3.6), we have by Lemma 3.1 (iii) that $\epsilon := H(\hat{\xi}) - H(\bar{\xi}) > 0$ (strict inequality), where $\hat{\xi}_{ij} = \max(0, r_i + r_j - \|\hat{c}_i - \hat{c}_j\|)$ and $\bar{\xi}$ is a solution of $P(\hat{c})$. Moreover, we can identify a neighborhood $\mathcal{N}$ of $\bar{c}$ such that for all $c^k \in \mathcal{N}$, we have

$$H(\xi^{k+1}) \leq H(\bar{\xi}^{k+1}) < H(\xi^k) - \epsilon/2, \tag{3.8}$$

This claim follows from Lemma 3.1 (iv) and the observation that the optimal objective in (3.6) is a continuous function of c^-, for c^- near $\hat{c}$. The fact that $\hat{c}_i \neq \hat{c}_j$ for all (i, j) ensures that the z_{ij} are continuous functions of c^-, while H itself is continuous by Assumption 1. Since there is a subsequence $\mathcal{S}$ with $\lim_{k \in \mathcal{S}} c^k = \hat{c}$, we have from (3.8) and monotonicity of the full sequence $\{H(\xi^k)\}$ that $H(\xi^k) \downarrow -\infty$. This is impossible, however, since H is bounded below by 0. We conclude therefore that all accumulation points $\hat{c}$ are either stationary or else have $\hat{c}_i = \hat{c}_j$ for some pair (i, j), as required. □

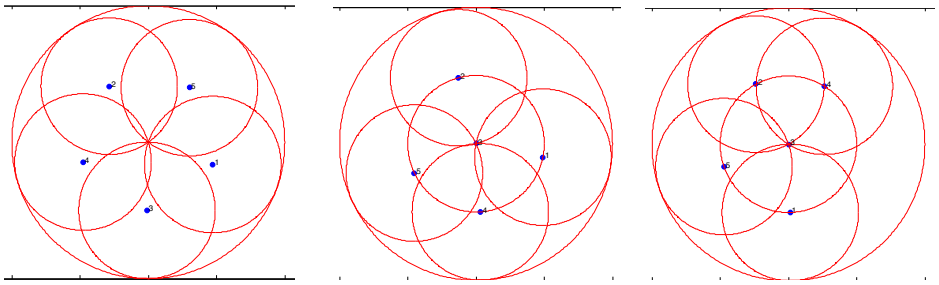
As noted above, the case in which accumulation points have coincident centers is exceptional, so Theorem 3.2 shows that the algorithm usually either terminates or accumulates at stationary points.

3.2. Examples. We present several examples showing results obtained with Algorithm 1 on various problems, and compare them with known results. To begin, a simple example demonstrates the existence of local minima that are not global minima.

EXAMPLE 3.1 (Five Circles). *Consider the problem of packing five circles of radius .5 into an enclosing circle of radius 1. Results obtained with Algorithm 1, with objective $H(\xi) = \|\xi\|_\infty$, from random starting points reveal an apparent global solution (Figure 3.1(a)) and a family of local solutions (Figures 3.1(b) and 3.1(c)). In the local solutions, one of the packed circles is positioned in the center of the enclosing circle and the remaining four circles are arranged around the boundary, in such a way that the maximum overlap between any pair of circles is .5. Algorithm 1 required only a few iterations for each of these examples.*

As noted in Section 1, optimal sphere packings (configurations with no overlap) have been obtained in two and three dimensions, for spaces of infinite extent. Our algorithm can solve only problems with finite enclosing shapes, but we can use large enclosures to investigate the similarities between local solutions attained by our algorithm and known optimal packings in $\mathbb{R}^2$ (hexagonal lattice with density $\pi/\sqrt{12}$) and $\mathbb{R}^3$ (FCC lattice with density $\pi/\sqrt{18}$).

EXAMPLE 3.2 (Uniform Circles in $\mathbb{R}^2$). *We ran Algorithm 1 with $N = 150$ circles, each of area π, and a circular container of size $150\sqrt{12}$. This results in a*



(a) Global Solution: $o = .4122147478$ (b) Local Solution: $o = .5$ (c) Local Solution: $o = .5$

FIG. 3.1. *Solutions obtained by Algorithm 1 for packing circles of radius* .5 *into a circle of radius* 1*, showing final overlap measures for each.*

total circle area-to-container area ratio which is equal to the optimal packing density. The resulting circle configuration is shown in Figure 3.2. The hexagonal arrangement of the circles is clearly visible in the interior of the container.

We also ran tests in which 100 circles are packed into a square container. (Rectangular feasible sets Ω_i are easily incorporated into the formulation by defining bound constraints on the centers c_i.) We generate starting points by arranging the centers in a 10×10 square lattice. We may then add a random perturbation to each center. Results are shown in Figure 3.3. (For clarity, we show only the centers in this figure, omitting the circles.) When no perturbations are added to the starting configuration, the algorithm does not move from the initial square configuration shown in Figure 3.3(a). When random initial perturbations are applied (large enough that the original square grid structure is not recognizable in the initial point), many different local minima are obtained. Three of these are shown in Figures 3.3(b), 3.3(c), and 3.3(d). Note that all of these have a maximum overlap less than the square configuration, and that hexagonal structure is recognizable in large parts of the domain, with square structure and disorder in intermediate regions.

EXAMPLE 3.3 (Uniform spheres in $\mathbb{R}^3$). *We performed a similar test to Example 3.2 in three dimensions. We checked to see whether Algorithm 1 converges to a solution like the FCC lattice in a finite minimum-overlap arrangement with 200 spheres enclosed in a larger sphere. We chose the small spheres to have volume π and the containing sphere to have volume $200\sqrt{18}$, giving a density of $\pi/\sqrt{18}$, identical to the FCC lattice, which is optimal in infinite space. At the solution obtained by Algo-*

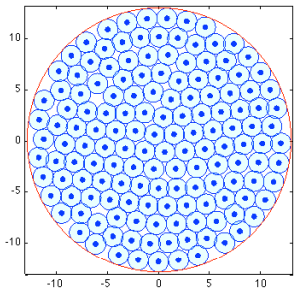
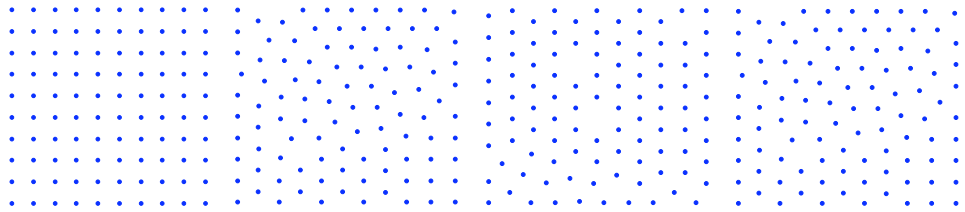


FIG. 3.2. *Circle packing in a circular enclosure. A nearly hexagonal arrangement is seen in the interior.*

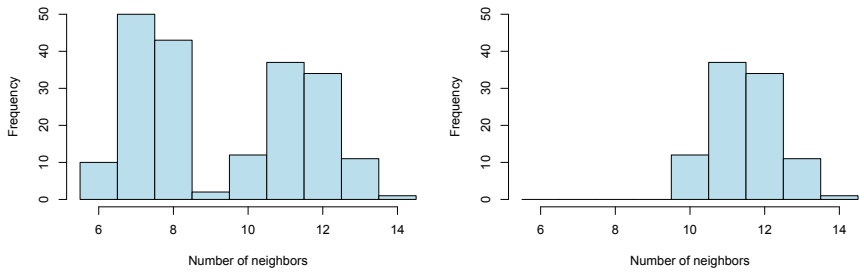


(a) $o = .1192514295$ (b) $o = .1188906843$ (c) $o = .1181440939$ (d) $o = .1179656050$

FIG. 3.3. *Local minima obtained by Algorithm 1 for packing circles into a square, showing final overlap measures for each.*

rithm 1, we counted the number of spheres that touch or intersect each sphere. This statistic provides an indication of the type of packing attained, since the FCC lattice has 12 neighbors per sphere, while the BCC lattice has only 8 neighbors per sphere. The histogram for the number of neighboring spheres is shown in Figure 3.4(a). A more instructive diagram is obtained by removing from consideration those spheres that touch the enclosing sphere. After doing so, we obtain the histogram in Figure 3.4(b). Each interior sphere has an average of about 11.5 contacts with its neighbors, suggesting strongly that the calculated solution is close to the FCC lattice in the interior of the domain.

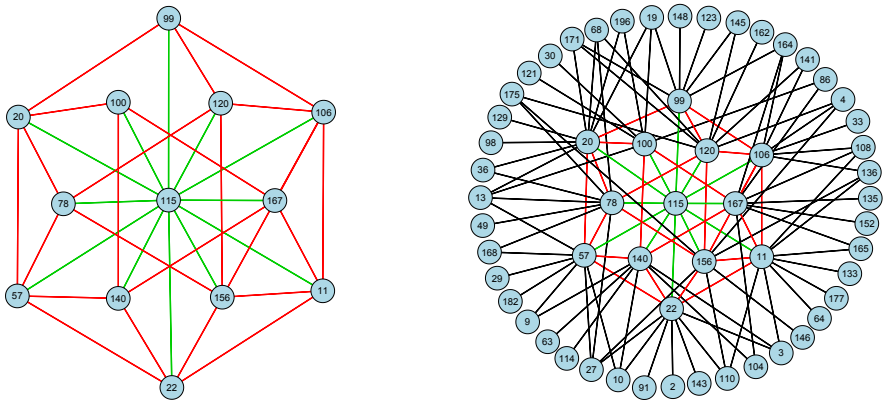
For further evidence, we constructed a graph where the vertices correspond to spheres and the edges represent overlap or "touching" with neighboring spheres. We chose one centrally located sphere and graphed its contacts with neighbors (see Figure 3.5(a)) and neighbors-of-neighbors (Figure 3.5(b)). These graphs differ in only minor ways from the graphs for the FCC arrangement. In comparison to Figure 3.5(a), the neighbor graph for the true FCC lattice for distance one would have one additional edge connecting sphere 99 to sphere 100 and no edge connecting sphere 156 with sphere 167. Our distance-two neighbor graph (Figure 3.5(b)) shows 44 spheres, exactly the same number as are present in the FCC lattice. We note finally that in the true FCC lattice, each first-level sphere has seven neighbors at the second level. Our arrangement shows an average of 6.75: Five spheres have six neighbors, five spheres have seven neighbors, and two spheres have eight neighbors. In summary, neighbor graph analysis shows further evidence that our 200-sphere arrangement is close to the FCC lattice, in the interior of the domain.



(a) Distribution of neighbor counts.

(b) Distribution of neighbor counts after removing spheres at the periphery.

FIG. 3.4. *Neighbor counts for packing of 100 three-dimensional spheres in in a spherical container.*



(a) Neighbor graph for distance 1

(b) Neighbor graph for distance 2

FIG. 3.5. *Graph representing the sphere arrangement for one central sphere and neighboring spheres up to a distance of 2. Vertices correspond to spheres and edges represent overlap or "touching".*

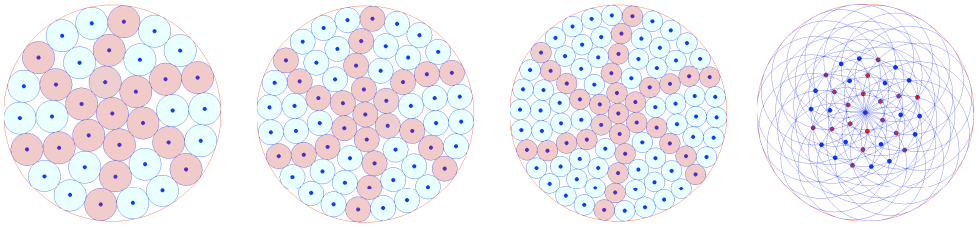
Finally, we report on solutions obtained by Algorithm 1 on packings of discs in a circle, for which the optimal packing is known in only a few cases. In particular, we analyze packings with equally sized discs, where the number of discs is given by a hexagonal number, that is,

$$h(k) = 3k(k+1) + 1, \qquad k \geq 1. \tag{3.9}$$

EXAMPLE 3.4. *Lubachevsky and Graham [13] introduce* curved hexagonal packings*, which are the best packings found so far for $h(k)$ defined by (3.9), for $k \leq 5$. We ran Algorithm 1 with the optimal densities found in [13] for $k = 3, 4, 5$. The best local optima we found are shown in Figure 3.6; they are identical to the configurations found in [13]. (We highlight some of the circles to emphasize the "curved" feature of the packing, which distinguishes it from a standard hexagonal arrangement, which has slightly lower density when restricted to a finite circle.)*

When we ran Algorithm 1 on the problem of 37 uniform discs in a larger disc, where the radii were too large to allow packing without overlap, the algorithm with $H(\xi) = \|\xi\|_\infty$ produced the same arrangement of centers as in Figure 3.6(a) (see Figure 3.6(d)) when initialized at a sufficiently close initial point. The maximal overlap is attained by many pairs of circles. We can obtain non-overlapping configurations by simply reducing the radii of all discs uniformly, by an amount equal to half the maximal overlap, to obtain a solution in which each pair of circles that formerly overlapped maximally now just touches.

4. Ellipsoid Packing. Here we discuss a bilevel optimization procedure for packing ellipsoids into an ellipsoidal container in a way that minimizes the maximum overlap of any pair of ellipsoids. It is not as obvious how to measure the overlap between two ellipsoids as between two spheres, since it depends on the orientation of the ellipsoids as well as the location of their centers. We measure the overlap by the *sum of principal semi-axes of the largest ellipsoid that can be inscribed in the intersection of the two ellipsoids.* This overlap measure can be calculated by solving a small semidefinite optimization problem, constructed according to the S-procedure (see Subsection 4.1). These are the lower-level problems in our bilevel optimization



(a) 37 discs (b) 61 discs (c) 91 discs (d) 37 discs, larger radii

FIG. 3.6. *Optimal configurations found using Algorithm 1 for hexagonal numbers of unit discs in an enclosing circle.*

formulation. The upper-level problem is to position and orient the ellipsoids so as to minimize the maximum overlap (see Subsection 4.2), while keeping all ellipsoids inside the enclosing shape. We refer to this problem as "min-max-overlap." Dual information from the lower-level problems provides a measure of sensitivity of the overlaps to the ellipsoid parameters, allowing us to develop a successive approximation approach, with trust regions, whose accumulation points are stationary for the min-max-overlap problem. Foundations for the convergence analysis of the trust-region algorithm are laid in Subsection 4.3, while the algorithm itself is outlined in Subsection 4.4, along with the final convergence result.

4.1. Measuring Overlap. Boyd and Vandenberghe [4, Section 8.4.2] consider the problem of finding the ellipsoid of largest volume inscribed in an intersection of ellipsoids. The volume of an ellipsoid $\mathcal{E} = \{c + Su \mid \|u\|_2 \leq 1\}$ is proportional to $\det(S)$. This problem is convex, and is best solved as a nonlinear semidefinite program (SDP), although a formulation as a linear SDP is also possible (see [17]). We consider an alternative in which the objective is taken to be $\mathrm{trace}(S)$, the sum of lengths of the semi-axes of the ellipsoid — a good proxy for the volume in problems of the type we consider. The trace admits a more straightforward linear SDP formulation than the maximum-determinant measure, thus facilitating theoretical development and analysis of our min-max-overlap problem and faster computations.

Recalling from (2.1) that we define the ellipsoid $\mathcal{E}_i$ by

$$\mathcal{E}_i := \{c_i + S_i u \mid \|u\|_2 \leq 1\}, \tag{4.1}$$

we introduce the notation $\Sigma_i = S_i^2$. Parametrizing the inscribed ellipsoid similarly by $\mathcal{E}_{ij} := \{c_{ij} + S_{ij} u \mid \|u\|_2 \leq 1\}$, and using (2.3) to formulate the fact that the inscribed ellipsoid is contained in both $\mathcal{E}_i$ and $\mathcal{E}_j$, we formulate the problem of measuring overlap as follows:

$$\begin{aligned}
\hat{O}(c_i, c_j, \Sigma_i, \Sigma_j) := \max_{S_{ij} \succeq 0, c_{ij}, \lambda_{ij1}, \lambda_{ij2}} \ & \mathrm{trace}(S_{ij}) \\
\text{subject to} \quad & \begin{pmatrix} -\lambda_{ij1} I & 0 & S_{ij} \\ 0 & \lambda_{ij1} - 1 & (c_{ij} - c_i)^T \\ S_{ij} & c_{ij} - c_i & -\Sigma_i \end{pmatrix} \preceq 0, \quad (4.2a) \\
& \begin{pmatrix} -\lambda_{ij2} I & 0 & S_{ij} \\ 0 & \lambda_{ij2} - 1 & (c_{ij} - c_j)^T \\ S_{ij} & c_{ij} - c_j & -\Sigma_j \end{pmatrix} \preceq 0. \quad (4.2b)
\end{aligned}$$

The Lagrangian can be written as

$$\begin{aligned}\mathcal{L}(c_{ij}, S_{ij}, \lambda_{ij1}, \lambda_{ij2}, T_{ij}, M_{ij1}, M_{ij2}) :=& \langle I, S_{ij}\rangle + \langle T_{ij}, S_{ij}\rangle \\ &- \langle M_{ij1}, \begin{pmatrix} -\lambda_{ij1} I & 0 & S_{ij} \\ 0 & \lambda_{ij1} - 1 & (c_{ij} - c_i)^T \\ S_{ij} & c_{ij} - c_i & -\Sigma_i \end{pmatrix}\rangle \\ &- \langle M_{ij2}, \begin{pmatrix} -\lambda_{ij2} I & 0 & S_{ij} \\ 0 & \lambda_{ij2} - 1 & (c_{ij} - c_j)^T \\ S_{ij} & c_{ij} - c_j & -\Sigma_j \end{pmatrix}\rangle,\end{aligned}$$

with the dual problem being derived from

$$\min_{M_{ij1}\succeq 0, M_{ij2}\succeq 0, T_{ij}\succeq 0} \left\{ \max_{S_{ij}\succeq 0, c_{ij}, \lambda_{ij1}, \lambda_{ij2}} \mathcal{L}(c_{ij}, S_{ij}, \lambda_{ij1}, \lambda_{ij2}, T_{ij}, M_{ij1}, M_{ij2}) \right\}.$$

Introducing the following notation for M_{ij1} and M_{ij2}:

$$M_{ij1} := \begin{pmatrix} R_{ij1} & r_{ij1} & P_{ij1} \\ r_{ij1}^T & p_{ij1} & q_{ij1}^T \\ P_{ij1} & q_{ij1} & Q_{ij1} \end{pmatrix}, \qquad M_{ij2} := \begin{pmatrix} R_{ij2} & r_{ij2} & P_{ij2} \\ r_{ij2}^T & p_{ij2} & q_{ij2}^T \\ P_{ij2} & q_{ij2} & Q_{ij2} \end{pmatrix}, \tag{4.3}$$

we can write the dual explicitly as follows:

$$\begin{aligned}\hat{O}(c_i, c_j, \Sigma_i, \Sigma_j) := \min_{M_{ij1}\succeq 0, M_{ij2}\succeq 0, T_{ij}\succeq 0} \quad & p_{ij1} + p_{ij2} + 2q_{ij1}^T c_i + 2q_{ij2}^T c_j \\ & + \langle Q_{ij1}, \Sigma_i\rangle + \langle Q_{ij2}, \Sigma_j\rangle \\ \text{subject to} \quad & 0 = I + T_{ij} - 2P_{ij1} - 2P_{ij2} \\ & 0 = \text{trace}(R_{ij1}) - p_{ij1} \\ & 0 = \text{trace}(R_{ij2}) - p_{ij2} \\ & 0 = q_{ij1} + q_{ij2}.\end{aligned} \tag{4.4}$$

(We have assumed without loss of generality that P_{ij1} and P_{ij2} are in $S\mathbb{R}^{n\times n}$; this follows from $S_{ij} \in S\mathbb{R}^{n\times n}$.)

A strong duality relationship holds between problems (4.2) and (4.4) because Slater's constraint qualification is satisfied for the second problem — it has a strictly feasible point. We construct this point by setting $T_{ij} = I$, and defining

$$M_{ij1} = M_{ij2} = \begin{pmatrix} I & 0 & \frac{1}{2}I \\ 0 & n & 0 \\ \frac{1}{2}I & 0 & I \end{pmatrix}.$$

It is easy to verify that these choices satisfy the linear constraints in (4.4), along with the (strict) interiority conditions $M_{ij1} \succ 0$, $M_{ij2} \succ 0$, $T_{ij} \succ 0$. The strong duality argument in Appendix B can now be applied to deduce that the optimal values of the problems (4.2) and (4.4) are identical, and are both $-\infty$ when (4.2) is infeasible (that is, the two ellipsoids in question do not touch). This fact justifies our use of the notation $\hat{O}(c_i, c_j, \Sigma_i, \Sigma_j)$ to denote the optimal objectives of both primal and dual.

4.2. Choosing Ellipsoid Positions and Orientations. The main variables in the min-max-overlap problem are the parameters defining the ellipsoids $\mathcal{E}_i$ for $i = 1, 2, \ldots, N$: the centers c_i and the orientations defined by S_i (and thus $\Sigma_i = S_i^2$). For $n = 3$ (which we assume in this section and subsequently), we would like to fix the lengths of the axes of each ellipsoid at the values r_{i1}, r_{i2}, and r_{i3} (assuming that $r_{i1} \geq r_{i2} \geq r_{i3}$). This is equivalent to fixing the eigenvalues of Σ_i at r_{i1}^2, r_{i2}^2, and r_{i3}^2, or to fixing the eigenvalues of S_i to r_{i1}, r_{i2}, and r_{i3}.

Using the notation $\hat{O}$ defined in (4.2) and (4.4), we can formulate the min-max-overlap problem as the following bilevel optimization problem:

$$\min_{\xi,(c_i,S_i,\Sigma_i),i=1,2,\ldots,N} \xi \tag{4.5a}$$

$$\text{subject to} \quad \xi \geq \hat{O}(c_i, c_j, \Sigma_i, \Sigma_j), \qquad 1 \leq i < j \leq N, \tag{4.5b}$$

$$\mathcal{E}_i \subset \mathcal{E}, \qquad i = 1, 2, \ldots, N, \tag{4.5c}$$

$$\Sigma_i = S_i^2, \qquad i = 1, 2, \ldots, N, \tag{4.5d}$$

$$\text{semi-axes of } \mathcal{E}_i \text{ have lengths } r_{i1}, r_{i2}, r_{i3}, \quad i = 1, 2, \ldots, N. \tag{4.5e}$$

This problem is nonconvex for three reasons. First, each pairwise overlap objective $\hat{O}(c_i, c_j, \Sigma_i, \Sigma_j)$ is a nonconvex function of its arguments. This issue is intrinsic; as in the sphere-packing problem, we expect there to be many local solutions and we can only expect our algorithm to find one of them. As we see below (in (4.10) and Algorithm 2), our algorithm iteratively solves subproblems in which each $\hat{O}$ is replaced by a linearized approximation that makes use of the optimal dual variables M_{ij1} and M_{ij2} from the formulation (4.4). These subproblems will be convex if we can overcome the other two sources of nonconvexity in the formulation (4.5), which we address now.

The second nonconvexity issue is in the constraint (4.5e) on the eigenvalues of S_i, $i = 1, 2, \ldots, N$. We consider instead the following convex relaxation:

$$S_i - r_{i1} I \preceq 0, \quad S_i - r_{i3} I \succeq 0, \quad \text{trace}(S_i) = r_{i1} + r_{i2} + r_{i3}. \tag{4.6}$$

Note that this is indeed a relaxation: If $\mathcal{E}_i$ has the desired dimensions, then the eigenvalues of S_i are r_{i1}, r_{i2}, and r_{i3}, and the conditions (4.6) are satisfied. Because the overall goal is to minimize maximal overlap, and because minimum-volume ellipsoids are those that are most eccentric, the relaxation (4.6) is usually "tight" with respect to (4.5) in many interesting cases. Intermediate iterates are often observed to have ellipsoids less eccentric than desired.

The third source of nonconvexity — the constraint (4.5d) — is relatively easy to deal with. We replace it with the following pair of restrictions:

$$\begin{bmatrix} \Sigma_i & S_i \\ S_i & I \end{bmatrix} \succeq 0, \qquad S_i \succeq 0, \qquad i = 1, 2, \ldots, N. \tag{4.7}$$

The first of these conditions ensures only that $\Sigma_i \succeq S_i^2$. However, the overlap $\hat{O}(c_i, c_j, \Sigma_i, \Sigma_j)$ will grow if Σ_i is replaced by any matrix $\tilde{\Sigma}_i \succeq \Sigma_i$. Hence, because of the min-max-overlap objective in (4.5), the matrices Σ_i will be set to the "smallest possible values" that satisfy the conditions (4.7), that is, $\Sigma_i = S_i^2$.

Finally, defining the containing ellipsoid to be $\mathcal{E} := \{x \mid (x - c)^T \Sigma^{-1} (x - c) \leq 1\}$, we can use (2.3) to formulate the condition (4.5c) as follows:

$$\begin{bmatrix} -\lambda_i I & 0 & S_i \\ 0 & \lambda_i - 1 & (c_i - c)^T \\ S_i & c_i - c & -\Sigma \end{bmatrix} \preceq 0, \tag{4.8}$$

Given all these considerations, we define the relaxed version of (4.5) to be addressed in this section as follows:

$$\min_{\xi,(\lambda_i,c_i,S_i,\Sigma_i),i=1,2,\ldots,N} \quad \xi \tag{4.9a}$$

$$\text{subject to} \quad \xi \geq \hat{O}(c_i, c_j, \Sigma_i, \Sigma_j), \qquad 1 \leq i < j \leq N, \tag{4.9b}$$

$$\begin{bmatrix} -\lambda_i I & 0 & S_i \\ 0 & \lambda_i - 1 & (c_i - c)^T \\ S_i & c_i - c & -\Sigma \end{bmatrix} \preceq 0, \quad i = 1, 2, \ldots, N, \tag{4.9c}$$

$$\begin{bmatrix} \Sigma_i & S_i \\ S_i & I \end{bmatrix} \succeq 0, \qquad i = 1, 2, \ldots, N, \tag{4.9d}$$

$$S_i - r_{i1} I \preceq 0, \quad S_i - r_{i3} I \succeq 0, \qquad i = 1, 2, \ldots, N, \tag{4.9e}$$

$$\text{trace}(S_i) = r_{i1} + r_{i2} + r_{i3}, \qquad i = 1, 2, \ldots, N. \tag{4.9f}$$

Note that when the ellipsoid $\mathcal{E}_i$ is actually a circle, that is, $r_{i1} = r_{i2} = r_{i3}$, we can fix $S_i = r_{i1} I$ and $\Sigma_i = r_{i1}^2 I$ in (4.10), and eliminate these variables from that problem. Hence, we can assume without loss of generality that $r_{i1} > r_{i3}$.

In the remainder of this section, we describe our algorithm for solving the bilevel optimization problem (4.9), and prove convergence properties. Our development and analysis takes place in a general setting that encompasses (4.9) but uses simpler notation. A key step in the algorithm is the solution of a subproblem for (4.9) in which the objective is linearized using the optimal values from the dual overlap formulation (4.4). The other constraints in (4.9) remain the same, and a trust region is added to restrict the amount by which the ellipsoid parameters can change. This subproblem can be stated as follows:

$$\min_{\xi,(\lambda_i,c_i,S_i,\Sigma_i),i=1,2,\ldots,N} \quad \xi \tag{4.10a}$$

$$\begin{aligned} \text{subject to} \quad & \xi \geq p_{ij1} + p_{ij2} + 2q_{ij1}^T c_i + 2q_{ij2}^T c_j \\ & \qquad + \langle Q_{ij1}, \Sigma_i \rangle + \langle Q_{ij2}, \Sigma_j \rangle, \qquad \text{for } (i,j) \in \mathcal{I}, \end{aligned} \tag{4.10b}$$

$$\begin{bmatrix} -\lambda_i I & 0 & S_i \\ 0 & \lambda_i - 1 & (c_i - c)^T \\ S_i & c_i - c & -\Sigma \end{bmatrix} \preceq 0, \quad i = 1, 2, \ldots, N, \tag{4.10c}$$

$$\begin{bmatrix} \Sigma_i & S_i \\ S_i & I \end{bmatrix} \succeq 0, \qquad i = 1, 2, \ldots, N, \tag{4.10d}$$

$$S_i - r_{i1} I \preceq 0, \quad S_i - r_{i3} I \succeq 0, \qquad i = 1, 2, \ldots, N, \tag{4.10e}$$

$$\text{trace}(S_i) = r_{i1} + r_{i2} + r_{i3}, \qquad i = 1, 2, \ldots, N, \tag{4.10f}$$

$$\|c_i - c_i^-\|_2^2 \leq \Delta_c^2, \qquad i = 1, 2, \ldots, N, \tag{4.10g}$$

$$\|S_i - S_i^-\| \leq \Delta_S, \qquad i = 1, 2, \ldots, N, \tag{4.10h}$$

$$|\lambda_i - \lambda_i^-| \leq \Delta_\lambda, \qquad i = 1, 2, \ldots, N. \tag{4.10i}$$

Here, $(\lambda_i^-, c_i^-, S_i^-, \Sigma_i^-)$, $i = 1, 2, \ldots, N$ are the values of the variables at the current iteration, and Δ_c, Δ_S, and Δ_λ are trust-region radii. The quantities p_{ij1}, p_{ij2}, q_{ij1}, q_{ij2}, Q_{ij1}, and Q_{ij2} are extracted from the dual solutions M_{ij1} and M_{ij2} of (4.4), evaluated at the current iterate's parameters c_i^-, c_j^-, Σ_i^-, Σ_j^-, according to the structure (4.3). The set $\mathcal{I}$ represents a *subset* of all possible pairs (i, j) for $1 \leq i < j \leq N$, representing some selection of ellipsoids which currently have a (strict) overlap. Further details on the choice of $\mathcal{I}$ are given in Subsection 4.4.

If it is possible to fit each ellipsoid $\mathcal{E}_i$ *strictly* inside the enclosing ellipsoid $\mathcal{E}$, the subproblem (4.10) satisfies a Slater condition. That is, there exists a point that satisfies the linear equality constraints and *strictly* satisfies the inequality constraints in this problem. We can prove this claim by observing that it is possible to find a point $(\tilde{\lambda}_i, \tilde{c}_i, \tilde{S}_i, \tilde{\Sigma}_i)$ that strictly satisfies the conditions (4.10c), (4.10d), (4.10e), along with the linear constraint (4.10f), and we construct the Slater point by taking a small step toward $(\tilde{\lambda}_i, \tilde{c}_i, \tilde{S}_i, \tilde{\Sigma}_i)$ from the current iterate $(\lambda_i^-, c_i^-, S_i^-, \Sigma_i^-)$.

4.3. Convergence Analysis. We prove in this subsection some technical results that provide the basis for convergence of the trust-region strategy. We start by considering the problem of measuring overlap of a single pair of ellipsoids, introducing simplified notation and showing that the solution of the dual formulation of this problem yields sensitivity of this measure with respect to the centers and orientations of the two ellipses. Second, we state the min-max-overlap problem (4.5) in the simplified notation, and define a reference problem in which the max is taken over only a subset of ellipsoid pairs. Third, we consider a linearized version of this reference problem, which is the subproblem solved in the trust-region algorithm. Finally, we state technical results that provide the foundation for the convergence analysis of the trust-region algorithm (Algorithm 2), which is presented in the next subsection. Several of the proofs are moved to Appendix C, to shorten the discussion.

4.3.1. Properties of the Overlap Measure of an Ellipsoid Pair. To simplify the notation, we note that each dual overlap problem (4.4) has the general form

$$P(l, C): \qquad t_l^*(C) := \min_{M_l} \langle C, M_l \rangle \tag{4.11a}$$

$$\text{s.t. } \langle A_{l,h}, M_l \rangle = b_{l,h}, \quad h = 1, 2, \ldots, p_l, \quad M_l \succeq 0. \tag{4.11b}$$

Here C captures the parameters that describe all the ellipsoids and M_l is the dual variable for the overlap problem (4.4). We construct C as a block-diagonal matrix with $N+1$ diagonal blocks. First, there are N diagonal blocks of the form

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & c_i^T \\ 0 & c_i & \Sigma_i \end{bmatrix}, \quad i = 1, 2, \ldots, N, \tag{4.12}$$

where each such block has dimension 7×7 and has the same partitioning scheme as the matrices M_{ij1} and M_{ij2} in (4.3). The remaining diagonal block in C is simply a 3×3 zero matrix that is used as the coefficient of the variable T_{ij} in (4.4), for each pair (i, j). The variable M_l in (4.11) has the same size and the same and block-diagonal structure as C, where M_{ij1} occupies the ith block-diagonal location, M_{ij2} occupies the jth block-diagonal location, and T_{ij} occupies the $(N+1)$st block-diagonal location (which is the 3×3 submatrix that appears in the lower right corner of M_l). In the constraints (4.11b), the matrices $A_{l,h}$ are chosen to capture the constraints in (4.4) (all of which are equalities). Structural constraints that enforce zeros in the locations of M_l not occupied by the M_{ij1}, M_{ij2}, and T_{ij} may be added to the formulation, but they are not necessary.

We show next that the constraints in (4.9) implicitly restrict C to a set Ω of the following form:

$$\Omega := \bar{\Omega} \cap \{C \,:\, \langle B_k, C \rangle = b_k, \; k = 1, 2, \ldots, p\}, \tag{4.13}$$

where $\bar{\Omega}$ is closed, convex, bounded, with nonempty interior. First, we show boundedness of the set of pairs (c_i, Σ_i) such that there exist λ_i and S_i for which conditions (4.9c), (4.9d), (4.9e), and (4.9f) are satisfied, recalling our assumptions that it is possible to place each ellipsoid $\mathcal{E}_i$ *strictly* inside the containing ellipsoid $\mathcal{E}$, and that the matrices Σ_i for spherical ellipsoids have been eliminated from the problem. Boundedness of c_i is immediately obvious from this containment condition. Boundedness of Σ_i is actually not guaranteed by the constraints (4.9c), (4.9d), (4.9e), and (4.9f), but we could add the constraint $\Sigma_i \preceq 2r_{i1}^2 I$ without changing the solution of the problem. (For simplicity, however, we do not apply this explicit bound to Σ_i in our discussion below.) Second, convexity and closedness of the set of pairs (c_i, Σ_i) for which there exist λ_i and S_i that satisfy (4.9c), (4.9d), and (4.9e) are immediate consequences of the form of these constraints. Third, we verify nonemptiness of the interior of the set of feasible (c_i, Σ_i) by taking the Slater point $(\tilde{\lambda}_i, \tilde{c}_i, \tilde{S}_i, \tilde{\Sigma}_i)$ discussed at the end of Subsection 4.2. Since at this point the matrices in (4.9c) and (4.9d) are negative and positive *definite*, respectively (rather than semidefinite), they will remain so under arbitrarily small perturbations to $\tilde{c}_i$ and $\tilde{\Sigma}_i$. Thus, $(\tilde{c}_i, \tilde{\Sigma}_i)$ is strictly interior to the set.

We now define the set $\bar{\Omega}$ to be the set of all symmetric matrices such that the elements in the locations corresponding to c_i and Σ_i are restricted by the constraints (4.9c), (4.9d), (4.9e), and (4.9f), restricting the elements in all other locations of the matrix to the interval $[0, 1]$. It is evident from the previous paragraph that the set $\bar{\Omega}$ is closed, convex, bounded, with nonempty interior.

It is now possible to impose the desired structure on the matrices C by using linear constraints (of the form $\langle B_k, C\rangle = b_k$ appearing in (4.13)). This structure includes the element 1 in the $(4, 4)$ position of (4.12), the zero elements in each diagonal block, and the zeros in all off-diagonal blocks.

Using notation analogous to that used in the discussion of the Slater condition at the end of Subsection 4.2, we denote by $\tilde{C}$ the point that satisfies

$$\tilde{C} \in \operatorname{int} \bar{\Omega} \quad \text{and} \quad \langle B_k, \tilde{C}\rangle = b_k, \;\; k = 1, 2, \ldots, p. \tag{4.14}$$

We denote by $M_l(C)$ an optimal value of M_l in (4.11) (not necessarily unique).

The primal form (4.2) of the overlap problem (4.11) has the form

$$\max_{\zeta_l = (\zeta_{l,1}, \zeta_{l,2}, \ldots, \zeta_{l,p_l})} b_l^T \zeta_l \quad \text{s.t.} \quad C - \sum_{h=1}^{p_l} \zeta_{l,h} A_{l,h} \succeq 0. \tag{4.15}$$

We noted in Subsection 4.1 that (4.4) (and thus (4.11)) always has a strictly feasible point. It follows from Corollary B.2 (in Appendix B) that strong duality holds, that is, the optimal values of (4.11) and (4.15) are identical. When there is a positive overlap between the two ellipsoids, the following optimality conditions relating the solutions of (4.11) and (4.15) are satisfied:

$$0 \preceq M_l \perp C - \sum_{h=1}^{p_l} \zeta_{l,h} A_{l,h} \succeq 0 \tag{4.16a}$$

$$\langle A_{l,h}, M_l\rangle = b_{l,h}, \;\; h = 1, 2, \ldots, p_l. \tag{4.16b}$$

In this case, it is evident from the nature of the problem that $t_l^*(C) > 0$. When the two ellipsoids touch without overlapping, we have $t_l^*(C) = 0$. When they do not even touch, the problem (4.15) is infeasible and (4.11) is unbounded, and we have

$t_l^*(C) = -\infty$. It is easy to see that $t_l^*(C)$ is a concave, extended-valued function of $C \in \Omega$, and as a consequence that $t_l^*(C)$ is continuous on the relative interior of its domain. Further useful facts about $t_l^*(C)$ are given in Lemma A.3. These include Lipschitz continuity in a neighborhood of a point C at which (4.15) has a strictly interior point (which, as noted in Subsection 4.1, occurs when the two ellipsoids have positive overlap), and the fact that any solution $M_l(C)$ of (4.11) belongs to the Clarke subdifferential of $t_l^*(C)$.

4.3.2. The Min-Max-Overlap Problem and a Reference Problem. Using the notation of (4.11) and (4.15) to capture the relaxed min-max-overlap problem (4.9), we can state this problem as follows:

$$P: \qquad \min_{C\in\Omega} \; t^*(C) := \max_{l=1,2,\ldots,m} t_l^*(C). \tag{4.17}$$

Here each element $l \in \{1, 2, \ldots, m\}$ represents the overlap problem for a single pair of ellipsoids. Note that $t^*(C) = -\infty$ if no pair of ellipsoids overlaps or touches.

We now define a similar problem that depends on just a subset $\mathcal{F} \subset \{1, 2, \ldots, m\}$ of the overlaps. The objective of this "reference problem" is

$$t_{\mathcal{F}}^*(C) := \max_{l\in\mathcal{F}} t_l^*(C), \tag{4.18}$$

where $\mathcal{F}$ is a subset of the *strictly overlapping* ellipsoid pairs, that is,

$$\mathcal{F} \subset \{l = 1, 2, \ldots, m \,:\, t_l^*(C) > 0\}.$$

(We will be more specific about the definition of $\mathcal{F}$ later.) By including the constraint $C \in \Omega$, we define the reference problem as follows:

$$P(\mathcal{F}): \qquad t_{\mathcal{F}}^* := \min_{C\in\Omega} t_{\mathcal{F}}^*(C) = \min_{C\in\Omega} \max_{l\in\mathcal{F}} t_l^*(C). \tag{4.19}$$

Nonsmooth analysis provides the following result that characterizes solutions of (4.19). A proof appears in Appendix C.

LEMMA 4.1. *Suppose that for a given set $\mathcal{F} \subset \{1, 2, \ldots, m\}$, $\bar{C}$ is a local solution of (4.19) at which (4.15) has a strictly interior point, for all $l \in \mathcal{F}$. Define $\bar{\mathcal{F}}(\bar{C})$ to be the set of indices achieving the maximum in (4.19), that is, $\bar{\mathcal{F}}(\bar{C}) = \{l \in \mathcal{F} : t_l^*(\bar{C}) = t_{\mathcal{F}}^*\}$. Then there exist $\bar{M}_l \in \partial t_l^*(\bar{C})$ and scalars μ_l, for all $l \in \bar{\mathcal{F}}(\bar{C})$, such that*

$$-\sum_{l\in\bar{\mathcal{F}}(\bar{C})} \mu_l \bar{M}_l \in N_\Omega(\bar{C}), \qquad \sum_{l\in\bar{\mathcal{F}}(\bar{C})} \mu_l = 1, \quad \mu_l \geq 0 \text{ for all } l \in \bar{\mathcal{F}}(\bar{C}), \quad \bar{C} \in \Omega. \tag{4.20}$$

By introducing multipliers for the indices $l \in \mathcal{F} \setminus \bar{\mathcal{F}}(\bar{C})$, we can restate the conditions (4.20) as follows:

$$0 \leq \mu_l \perp t_{\mathcal{F}}^* - t_l^*(\bar{C}) \geq 0, \;\; \text{for all } l \in \mathcal{F}, \tag{4.21a}$$

$$\sum_{l\in\mathcal{F}} \mu_l = 1, \tag{4.21b}$$

$$-\sum_{l\in\mathcal{F}} \mu_l \bar{M}_l \in N_\Omega(\bar{C}), \tag{4.21c}$$

$$\bar{C} \in \Omega. \tag{4.21d}$$

We say that a point $\bar{C}$ at which these conditions are satisfied is *Clarke-stationary* for $P(\mathcal{F})$ defined in (4.19).

4.3.3. Linearization of the Reference Problem. In the trust-region algorithm to be described below, the solutions $M_l(C)$ of (4.11) for $l \in \mathcal{F}$ are used to construct a linearized subproblem whose solution is a step ΔC in the parameter C, assuming that the current C is feasible. The subproblem is as follows:

$$L(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho): \qquad r(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho) := \min_{r, \Delta C} r \tag{4.22a}$$

$$\text{s.t.} \qquad r \geq t_l^*(C) + \langle \Delta C, M_l(C) \rangle, \;\; l \in \mathcal{F}, \tag{4.22b}$$

$$C + \Delta C \in \Omega, \quad \|\Delta C\| \leq \rho. \tag{4.22c}$$

Here $\rho > 0$ is a trust-region radius, and $M_{\mathcal{F}}(C)$ denotes the set of matrices $\{M_l(C) : l \in \mathcal{F}\}$. The problem (4.22) is convex, and its feasible set is bounded, so it has an optimal value which we denote by $\Delta C(\rho)$. Further, the KKT conditions are satisfied at this point. This claim follows from the fact that, given the point $\tilde{C}$ satisfying (4.14), and defining $\Delta C = \epsilon(\tilde{C} - C)$ for some small positive $\epsilon > 0$, we have that

$$C + \Delta C = (1 - \epsilon)C + \epsilon \tilde{C} \in \operatorname{int} \bar{\Omega},$$

while the trust-region condition is strictly satisfied ($\|\epsilon(\tilde{C} - C)\| < \rho$), and the remaining constraints in (4.22) are affine. Hence, the conditions of [19, Theorem 28.2] are satisfied, and we can apply [19, Corollary 28.3.1] to deduce that there exist μ_l, $l \in \mathcal{F}$ and $\tau \geq 0$ such that

$$1 - \sum_{l \in \mathcal{F}} \mu_l = 0, \tag{4.23a}$$

$$0 \leq \mu_l \perp r(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho) - t_l^*(C) - \langle \Delta C, M_l(C) \rangle \geq 0, \;\; l \in \mathcal{F}, \tag{4.23b}$$

$$-\sum_{l \in \mathcal{F}} \mu_l M_l(C) - \tau u \in N_\Omega(C + \Delta C) \;\; \text{for some } u \in \partial \|\Delta C\|, \tag{4.23c}$$

$$C + \Delta C \in \Omega, \tag{4.23d}$$

$$0 \leq \tau \perp \rho - \|\Delta C\| \geq 0. \tag{4.23e}$$

Here $N_\Omega(C)$ denotes the normal cone to the closed convex set Ω at the point C (see (1.1)) and ∂ denotes a subdifferential. (As noted in Section 1, since $\|\cdot\|$ is convex and Lipschitz continuous, the Clarke subdifferential coincides with the subdifferential from convex analysis.) Note that the set $\{\tau v : \tau \geq 0, \; v \in \partial \|\Delta C\|\}$ is a closed convex cone and that it is an outer semicontinuous set-valued function of ΔC.

It is convenient for the following discussion to define the function

$$G_{\mathcal{F}}(\Delta C; C, M_{\mathcal{F}}(C)) := \max_{l \in \mathcal{F}} \langle C + \Delta C, M_l(C) \rangle, \tag{4.24}$$

and rewrite $L(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho)$ in terms of $G_{\mathcal{F}}$ as follows:

$$L(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho): \;\; \min_{\Delta C} \; G_{\mathcal{F}}(\Delta C; C, M_{\mathcal{F}}(C)) \;\; \text{s.t.} \;\; C + \Delta C \in \Omega, \;\; \|\Delta C\| \leq \rho. \tag{4.25}$$

Note that $G_{\mathcal{F}}(\cdot; C, M_{\mathcal{F}}(C))$ is piecewise linear and convex.

4.3.4. Properties of the Linearized Reference Problem. For purposes of our main technical lemma, we define the "predicted decrease" from subproblem $L(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho)$ as follows:

$$\Lambda(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho) := t_{\mathcal{F}}^*(C) - r(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho). \tag{4.26}$$

Note that since $\Delta C = 0$ is feasible for (4.22), we have $\Lambda(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho) \geq 0$.

LEMMA 4.2. *Let $\mathcal{F} \subset \{1, 2, \ldots, m\}$ be given.*

(i) *Suppose that $\bar{C}$ is such that (4.15) has a strictly feasible point for all $l \in \mathcal{F}$. If $\Lambda(\mathcal{F}, \bar{C}, M_{\mathcal{F}}(\bar{C}), \rho) = 0$ for some $\rho > 0$ and some set of solutions $M_l(\bar{C})$ to (4.11) for $l \in \mathcal{F}$, then $\bar{C}$ is Clarke-stationary for $P(\mathcal{F})$.*

(ii) *$\Lambda(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho)$ is an increasing function of $\rho > 0$.*

(iii) *$\Lambda(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho)/\rho$ is a decreasing function of $\rho > 0$.*

(iv) *For all C in some neighborhood of $\bar{C}$ defined in (i), we have that $t^*_{\mathcal{F}}(C + \Delta C(\rho)) \leq r(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho)$ for any $\rho > 0$.*

A proof appears in Appendix C.

We show now that all accumulation points of a sequence $\{C_k\}$ for which

$$\Lambda(\mathcal{F}, C_k, M_{\mathcal{F}}(C_k), 1) \to 0$$

are Clarke-stationary for $P(\mathcal{F})$.

THEOREM 4.3. *Suppose that for a given set $\mathcal{F} \subset \{1, 2, \ldots, m\}$, $\{C_k\}$ is a sequence of matrices in Ω converging to a limit $\bar{C}$ such that (4.15) has a strictly feasible point for each $l \in \mathcal{F}$. Suppose further that $\Lambda(\mathcal{F}, C_k, M_{\mathcal{F}}(C_k), 1) \to 0$. Then $\bar{C}$ is Clarke-stationary for $P(\mathcal{F})$.*

Proof. We invoke Theorem A.2 to deduce that for all $l \in \mathcal{F}$, the solution sets of $P(l, C_k)$ in (4.11) are uniformly bounded for all k sufficiently large. Hence, we can identify subsequences of $\{M_l(C_k)\}$ for each $l \in \mathcal{F}$ that approach some limit $\bar{M}_l$, where by Theorem A.2 (ii), $\bar{M}_l$ is a solution of $P(l, \bar{C})$ for each $l \in \mathcal{F}$. We can thus write $M_l(\bar{C}) = \bar{M}_l$ for each $l \in \mathcal{F}$. Thus, by taking a subsequence, we have that $M_{\mathcal{F}}(C_k) \to M_{\mathcal{F}}(\bar{C})$.

We show next, by contradiction, that $\Lambda(\mathcal{F}, \bar{C}, M_{\mathcal{F}}(\bar{C}), 1) = 0$. If this claim is not true, there exists $\Delta C'$ such that

$$\|\Delta C'\| \leq 1, \quad \bar{C} + \Delta C' \in \Omega, \quad G_{\mathcal{F}}(\Delta C'; \bar{C}, M_{\mathcal{F}}(\bar{C})) \leq t^*_{\mathcal{F}}(\bar{C}) - \epsilon,$$

for some $\epsilon > 0$. Defining

$$\Delta C'_k := \bar{C} - C_k + \Delta C',$$

we have from $C_k \to \bar{C}$, $\|\Delta C_k\| \leq 1$, and $C_k + \Delta C'_k = \bar{C} + \Delta C' \in \Omega$ that $\Delta C'_k$ is feasible for $L(\mathcal{F}, C_k, M_{\mathcal{F}}(C_k), 2)$. Hence, invoking Lemma 4.2 (iii), we have

$$\begin{aligned}
\lim_k G_{\mathcal{F}}(\Delta C'_k; C_k, M_{\mathcal{F}}(C_k)) &\geq \lim_k t^*_{\mathcal{F}}(C_k) - \Lambda(\mathcal{F}, C_k, M_{\mathcal{F}}(C_k), 2) \\
&\geq \lim_k t^*_{\mathcal{F}}(C_k) - 2\Lambda(\mathcal{F}, C_k, M_{\mathcal{F}}(C_k), 1) \\
&= t^*_{\mathcal{F}}(\bar{C}).
\end{aligned}$$

The final limit above follows from the definition of $t^*_{\mathcal{F}}$, Lemma A.3 (iv), and the assumption that $\Lambda(\mathcal{F}, C_k, M_{\mathcal{F}}(C_k), 1) \to 0$. On the other hand, we have from $C_k + \Delta C'_k = \bar{C} + \Delta C'$, the definition of $G_{\mathcal{F}}$ (4.24), and the limit $M_{\mathcal{F}}(C_k) \to M_{\mathcal{F}}(\bar{C})$ that

$$\begin{aligned}
\lim_k G_{\mathcal{F}}(\Delta C'_k; C_k, M_{\mathcal{F}}(C_k)) &= \lim_k \max_{l \in \mathcal{F}} \langle C_k + \Delta C'_k, M_l(C_k) \rangle \\
&= \lim_k \max_{l \in \mathcal{F}} \langle \bar{C} + \Delta C', M_l(C_k) \rangle \\
&= \max_{l \in \mathcal{F}} \langle \bar{C} + \Delta C', M_l(\bar{C}) \rangle \\
&\leq t^*_{\mathcal{F}}(\bar{C}) - \epsilon,
\end{aligned}$$

where $\epsilon > 0$ is defined above. This yields the contradiction, so we conclude that $\Lambda(\mathcal{F}, \bar{C}, M_{\mathcal{F}}(\bar{C}, 1) = 0$, as claimed. Clarke stationarity of $\bar{C}$ for $P(\mathcal{F})$ now follows from Lemma 4.2 (i). □

4.4. Trust-Region Algorithm. We now define the algorithm for solving the problem P defined by (4.17). Note that in this general setting, $t_l^*(C)$ defined by (4.11) is continuous on the set

$$\Psi_l := \{C : t_l^*(C) > -\infty\},$$

which is closed and convex. We make additional assumptions about the nature of the solutions to the parametrized primal-dual pair (4.11), (4.15), that do not hold in general, but which are satisfied for the application we consider here.

ASSUMPTION 2.

(i) $t_l^*(C) > -\infty \;\Rightarrow\; t_l^*(C) \geq 0$.

(ii) If $t_l^*(C) > 0$, *the problem (4.15) has a strictly feasible point.*

It is an immediate consequence of Assumption 2 and Lemma A.3 that all points C on the boundary of Ψ_l have $t_l^*(C) = 0$. We also have the following uniform continuity result, for which the proof is again deferred to Appendix C.

LEMMA 4.4. *Suppose that* $t_l^*(C)$ *is defined by (4.11), that Assumption 2 holds, and that* Ω *has the form (4.13). Let* $\bar{t} > 0$ *be given. Then for any* $\epsilon > 0$, *there is* $\delta > 0$ *such that for all* $\bar{C} \in \Omega$, *all* $C \in \Omega$ *with* $\|C - \bar{C}\| \leq \delta$, *and all* $l = 1, 2, \ldots, m$, *the following conditions hold:*

(i) If $t_l^*(\bar{C}) \geq \bar{t}$, *then* $t_l^*(C) \geq t_l^*(\bar{C}) - \epsilon$;

(ii) $t_l^*(C) \leq \max(0, t_l^*(\bar{C})) + \epsilon$.

A key issue in implementing the algorithm is to decide which subset $\mathcal{F}$ of the overlapping pairs to use in calculating the step ΔC in (4.22). Clearly, $\mathcal{F}$ should include the indices l for which the overlaps between the corresponding ellipsoid pairs are at or near the maximum. It could also include other indices with positive (but smaller) overlap. Clearly, it cannot contain any non-overlapping ellipsoids, as the problem $P(l, C)$ (4.11) has no solution in this case, so $M_l(C)$ is not defined. We settle on the following requirement, which depends on parameters $\eta_1, \eta_2 \in (0, 1)$ with $0 < \eta_1 < \eta_2 < 1$: Given C_k for which $t^*(C_k) > 0$ (see definition (4.17)), we choose $\mathcal{F}_k$ to satisfy:

$$\{l : t_l^*(C_k) \geq \eta_2 t^*(C_k)\} \subset \mathcal{F}_k \subset \{l : t_l^*(C_k) \geq \eta_1 t^*(C_k)\}. \tag{4.27}$$

Algorithm 2 describes our method. It follows a standard trust-region framework, though its analysis is a little non-standard. At each iteration, we calculate a candidate step ΔC_k by solving the linearized subproblem (4.22) with trust-region radius ρ_k, and calculate the predicted reduction $\Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho_k)$ (4.26) expected from this step. If the actual objective achieves at least a fraction c_1 of this decrease (for $c_1 \in (0, 1)$), we accept the step. If in fact the improvement is at least a larger fraction c_2 of the expected decrease, we may increase the trust-region radius for the next iteration. Otherwise, we do not take the step, but rather shrink the trust-region radius and proceed to the next iteration.

We now show that when the values $t^*(C_k)$ are bounded away from zero, there is a positive threshold such that any step ΔC_k with norm smaller than this threshold will be accepted.

LEMMA 4.5. *Suppose that Assumption 2 holds and let* $\bar{t} > 0$ *be given. Let* C_k *be any iterate with* $t^*(C_k) \geq \bar{t}$ *such that* C_k *is not Clarke-stationary for the problem P defined in (4.17), and suppose that* $\mathcal{F}_k$ *satisfies (4.27). Then there exists a threshold value* $\bar{\rho}_{\bar{t}} > 0$ *(independent of* C_k*) such that*

$$t^*(C_k + \Delta C(\rho)) \leq t^*(C_k) - \Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho) < t^*(C_k) \tag{4.28}$$

Algorithm 2 Packing Ellipsoids by Minimizing Overlap

Given $\Omega \subset S\mathbb{R}^{n\times n}$ compact; $\eta \in (0,1)$; c_1 and c_2 with $0 < c_1 < c_2 < 1$; ϕ_1 and ϕ_2 with $0 < \phi_1 < 1 < \phi_2$; and $\rho_{\max} > 0$;
Choose $C_0 \in \Omega$, $\rho_0 \in (0, \rho_{\max}]$;
for $k = 0, 1, 2, \ldots$ **do**
  Define $\mathcal{F}_k$ as in (4.27);
  Solve $L(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho_k)$ (4.22) to obtain ΔC_k;
  Compute predicted decrease $\Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho_k)$ from (4.26);
  if $\Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho_k) = 0$ **then**
    stop and return C_k;
  end if
  if $t^*(C_k + \Delta C_k) \le t^*(C_k) - c_1\Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho_k)$ **then**
    $C_{k+1} \leftarrow C_k + \Delta C_k$;
    if $t^*(C_k + \Delta C_k) \le t^*(C_k) - c_2\Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho_k)$ **then**
      $\rho_{k+1} \leftarrow \min(\phi_2\rho_k, \rho_{\max})$;
    end if
    if $t^*(C_{k+1}) \le 0$ **then**
      stop and return C_{k+1};
    end if
  else
    $C_{k+1} \leftarrow C_k$;
    $\rho_{k+1} \leftarrow \phi_1\rho_k$;
  end if
end for

whenever $\|\Delta C(\rho)\|$ *is a solution of* $L(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho)$ *with* $\rho \in (0, \bar{\rho}_{\bar{t}}]$.

Proof. Note first that if C_k were Clarke-stationary for $P(\mathcal{F}_k)$, given that $\mathcal{F}_k$ contains all the indices l for which $t_l^*(C_k)$ attains the maximum $t^*(C_k)$, we would have that C_k is also Clarke-stationary for P, which we have assumed is not the case. From Assumption 2 and Lemma 4.2 (i) we have therefore that $\Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho) > 0$ for all $\rho > 0$ and all solutions $M_l(C_k)$ to (4.11) with $C = C_k$ and $l \in \mathcal{F}_k$.

Now define $\epsilon = \bar{t}(1-\eta_2)/2$, and let $\bar{\rho}_{\bar{t}}$ be the corresponding (positive) value of δ from Lemma 4.4. Consider any ΔC such that $\|\Delta C\| \le \bar{\rho}_{\bar{t}}$. For indices l such that $t_l^*(C_k) = t^*(C_k) \ge \bar{t}$, we have from Lemma 4.4 (i) that

$$t_l^*(C_k + \Delta C) \ge t^*(C_k) - \bar{t}(1-\eta_2)/2 \ge t^*(C_k)(1+\eta_2)/2.$$

For indices $l \notin \mathcal{F}_k$, we have $t_l^*(C_k) \le \eta_2 t^*(C_k)$, and so from Lemma 4.4 it follows that

$$t_l^*(C_k+\Delta C) \le \max(0, t_l^*(C_k)) + \bar{t}(1-\eta_2)/2 \le \eta_2 t^*(C_k) + \bar{t}(1-\eta_2)/2 \le t^*(C_k)(1+\eta_2)/2.$$

Hence, for $\|\Delta C\| \le \bar{\rho}_{\bar{t}}$, the index l for which $t^*(C_k + \Delta C) = t_l^*(C_k + \Delta C)$ comes from the set $\mathcal{F}_k$, that is,

$$t_{\mathcal{F}_k}^*(C_k + \Delta C) = t^*(C_k + \Delta C).$$

Thus, by applying Lemma 4.2 (iv) and choosing $\rho \in (0, \bar{\rho}_{\bar{t}}]$ and setting $\Delta C(\rho)$ to be

the solution of $L(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho)$, we find that

$$\begin{aligned} t^*(C_k + \Delta C(\rho)) &= t^*_{\mathcal{F}_k}(C_k + \Delta C(\rho)) \\ &\leq t^*_{\mathcal{F}_k}(C_k) - \Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho) \\ &< t^*_{\mathcal{F}_k}(C_k) \\ &= t^*(C_k), \end{aligned}$$

as claimed. □

The inequality (4.28) satisfies the step acceptance conditions in Algorithm 2, since $0 < c_1 < c_2 < 1$. It follows immediately that for any C_k with $t^*(C_k) > 0$, the algorithm cannot "get stuck" by performing infinitely many unsuccessful iterations — eventually it will decrease ρ to the point where the step acceptance condition holds.

We now prove the main convergence result.

THEOREM 4.6. *Suppose that Assumption 2 holds. Then either (a) Algorithm 2 terminates finitely at a point that is Clarke-stationary for problem P (4.17), or has a non-positive value of t^*; or (b) it generates an infinite sequence of iterates $\{C_k\}$ for which accumulation points exist, and all accumulation points $\bar{C}$ either are Clarke-stationary for P or have $t^*(\bar{C}) = 0$.*

Proof. Consider first case (a). The second option follows immediately from the second termination condition $t^*(C_{k+1}) \leq 0$. For the first termination condition, we have by Assumption 2 (ii) and Lemma 4.2 (i) that C_k is stationary for problem $P(\mathcal{F}_k)$, and thus for problem P, which covers the first option in case (a).

We focus now on case (b), in which the sequence $\{C_k\}$ is infinite. Since all iterates are confined to the compact set Ω, accumulation points of sequence $\{C_k\}$ exist. Since the sequence of function values $\{t^*(C_k)\}$ is decreasing, then we have $t^*(\bar{C}) = 0$ for an accumulation point $\bar{C}$ if and only if $t^*(C_k) \downarrow 0$, which is the second option in case (b). Otherwise, we must be able to define $\bar{t} > 0$ such that $t^*(C_k) \geq \bar{t}$ for all k. We complete the proof by showing that in this case, all accumulation points $\bar{C}$ are Clarke-stationary.

From Lemma 4.5, we see that at each iteration k, the trust-region radius ρ_k will generate a successful step ΔC_k whenever it falls below $\bar{\rho}_{\bar{t}}$. Hence, since the algorithm decreases ρ by a factor of ϕ_1 after each unsuccessful step, we have that

$$\rho_k \geq \min(\rho_0, \phi_1 \bar{\rho}_{\bar{t}}), \quad \text{for all } k. \tag{4.29}$$

In considering accumulation points of the sequence $\{C_k\}$ we can remove all repeated entries from the sequence. These repeats arise from unsuccessful steps (for which the acceptance condition was not satisfied), and the accumulation points of the sequence are the same whether the repeated entries are present or not. Note that there must be infinitely many successful steps since, as we note in the comment after Lemma 4.5, the algorithm must eventually move away from any non-stationary point C_k with $t^*(C_k) > 0$. We denote the subsequence of successful iterates by $\mathcal{S}$.

At a successful iteration $k \in \mathcal{S}$, we have

$$\begin{aligned} t^*(C_{k+1}) &= t^*(C_k + \Delta C_k) \\ &\leq t^*(C_k) - c_1 \Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho_k) \\ &\leq t^*(C_k) - c_1 \min(\rho_k, 1) \Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), 1) \\ &\leq t^*(C_k) - c_1 \min(\rho_0, \phi_1 \bar{\rho}_{\bar{t}}, 1) \Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), 1), \end{aligned}$$

where we used Lemma 4.2 (ii) and (iii) to derive the second inequality, and the last inequality comes from (4.29). Since the sequence $\{t^*(C_k)\}$ is decreasing and bounded below (by $\bar{t}$), we have $0 < t^*(C_k) - t^*(C_{k+1}) \downarrow 0$, so by rearranging and using the fact that $\min(\rho_0, \phi_1\bar{\rho}_{\bar{t}}, 1) > 0$, we have

$$\lim_{k\in\mathcal{S}} \Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), 1) = 0. \tag{4.30}$$

Now suppose that $\bar{C} \in \Omega$ is an accumulation point of the full sequence $\{C_k\}$. As noted above, it must also be an accumulation point of the "successful iterate" sequence $\{C_k\}_{k\in\mathcal{S}}$. So by taking a further subsequence $\mathcal{S}' \subset \mathcal{S}$, we have $\lim_{k\in\mathcal{S}'} C_k = \bar{C}$. Since there are only a finite number of possibilities for the set $\mathcal{F}_k$, we can take another subsequence $\mathcal{S}'' \subset \mathcal{S}'$ such that, in addition, $\mathcal{F}_k \equiv \mathcal{F}$ for all $k \in \mathcal{S}''$. By the definition (4.27), we have that $t_l^*(C_k) \geq \eta_1 t^*(C_k) \geq \eta_1\bar{t}$ for all $l \in \mathcal{F}_k$ and all $k \in \mathcal{S}''$. Thus, using continuity of t_l^*, we have that

$$t_l^*(\bar{C}) \geq \eta_1\bar{t} > 0, \qquad \text{for all } l \in \mathcal{F},$$

implying that $P(l, \bar{C})$ has a strictly feasible point for all $l \in \mathcal{F}$. Clarke stationarity of $\bar{C}$ for problem $P(\mathcal{F})$ now follows from (4.30), using the fact that $\mathcal{S}'' \subset \mathcal{S}$ and applying Theorem 4.3. Since $\mathcal{F}$ contains all indices l for which t_l^* attains its maximum, $\bar{C}$ is also Clarke-stationary for P. □

5. Application: Chromosomal Arrangement in Human Cell Nuclei. We return to the problem of finding arrangements of chromosome territories in a cell nucleus on the basis of simple geometric principles (namely, low overlap and discouragement of proximity for homologous pairs) and seeing how closely the resulting arrangements match the experimental observations that have been made to date.

During most of the cell cycle, the chromosomes of higher eukaryotes are organized into distinct compartments known as chromosome territories (CTs). These domains have a roughly ellipsoidal shape (see for example [12]) and can overlap each other. This overlap is believed to have an important biological purpose, since it allows for the interaction and co-regulation of different genes. Additionally, the CTs tend to exploit the space available inside the cell nucleus, to allow for internal DNA-free channels, the *interchromatin compartments*. These compartments allow CTs deep inside the cell nucleus to be accessible for regulatory factors; see the CT-IC (chromosome territory-interchromatin compartment) model in [6].

As noted in Section 1, the arrangement of CTs is known to be non-random. Arrangements are known to be broadly conserved during evolution and are similar among cell types with similar developmental pathways. CT arrangements can also change during processes such as cancer development or cell differentiation. See [6] for an overview on what is known about CT arrangements and [14, 21, 15] for evidence and more details about the non-randomness of CT positioning.

There is strong evidence that chromosomes have a preferred radial position inside the nucleus [1, 6]. These preferences appear to be different for nuclei of different shapes, spherical and ellipsoidal. In ellipsoidal nuclei, CT size seems to drive the radial preferences, with the smaller CTs tending to lie nearer to the center. In spherical nuclei the situation is less clear, with the more gene-dense chromosomes seeming to lie nearer to the center of the nucleus.

There is also evidence for neighbor preferences, which may play an important role in causing co-regulated genes in different chromosomes to be closer together. In particular, it has been observed recently that CTs tend to favor neighborhoods of

heterologous chromosomes [11, 12]. This results in the two chromosomes in a homologous pair tending to be well separated. (In human cells there are 22 homologous pairs, each consisting of one chromosome from the mother and a similar one from the father.)

We model a CT arrangement as a packing of overlapping ellipsoids of various sizes inside an ellipsoidal or spherical container representing the cell nucleus. Minimizing maximum overlap mimics the fact that the CTs exploit the space available in the nucleus, to allow for the presence of contiguous DNA-free interchromatin channels. In this section we analyze whether purely geometric considerations, enforcing the simple principles of minimal overlap and well-separatedness of homologous pairs, can explain the observed arrangements of CTs in cell nuclei of different sizes and shapes.

5.1. Human Cell Nucleus. The human cell nucleus has a volume of between 500 μm^3 and 1600 μm^3, depending on the cell size and stage of differentiation. The shape also differs according to cell type. Human fibroblasts, for example, have flat ellipsoidal nuclei, whereas lymphocytes have spherical nuclei. In this study we analyze three different nucleus sizes: small (500 μm^3), medium (1000 μm^3) and large (1600 μm^3). For all three sizes we consider two shapes: spherical nuclei and flat ellipsoidal nuclei (with axis lengths in the ratio 1:2:4).

We estimate the volume of each CT to be proportional to the number of base pairs in the chromosome, with the constant of proportionality determined by the average chromatin packing density. The number of base pairs for human cells ranges from 47 Mbp (chromosome 21) to 247 Mbp (chromosome 1), while the human chromatin packing density in living cells has been estimated to be 0.15 μm^3/Mbp ([16]). By multiplying the total number of base pairs by this average density, we arrive at a total volume of about 461 μm^3 over all CTs. The individual volumes for each chromosome territory are given in Table 5.1.

TABLE 5.1
Volume of each chromosome territory based on chromatin packing density of 0.15 μm^3*/Mbp.*

CT	1	2	3	4	5	6	7	8
volume	37.05	36.45	29.85	28.65	27.15	25.65	23.85	21.90
CT	9	10	11	12	13	14	15	16
volume	21.00	20.25	20.10	19.80	17.10	15.90	15.00	13.35
CT	17	18	19	20	21	22	X	Y
volume	11.85	11.40	9.45	9.30	7.05	7.50	23.25	8.70

5.2. Implementation. Algorithm 2 was implemented in Matlab and CVX [9]. It is terminated when one of the following conditions holds.

(i) The ratio of predicted decrease to trust-region radius falls below a specified tolerance. Using the notation of Algorithm 2, we state this condition as

$$\Lambda(\mathcal{F}_k, C_k, M_{\mathcal{F}_k}(C_k), \rho_k)/\rho_k \leq \texttt{Tol1},$$

where we set `Tol1` = .005 in our experiments.

(ii) The maximum overlap falls below a small fraction `Tol2` of the volume of the enclosing ellipsoid. We used `Tol2` = .0001 in our experiments.

(iii) The algorithm runs for 100 iterations.

Many instances of the problem, including problems of different sizes and shapes with different random starting points, were executed on servers running various versions of Linux.

5.3. Radial Preferences. In the first set of experiments, we use Algorithm 2 to solve the min-max-overlap problem for the CTs, enforcing no constraints or penalties on the homologous pairs. We set up numerous trials, with the data varied as follows.

(i) CT volumes are obtained by sampling from a normal distribution, with mean taken from Table 5.1 and the standard deviation set to .02 of the mean.

(ii) The relative axis lengths are varied around the intercepts found in [12] for mouse chromosomes, namely, 1:2.9:4.4. The second and third ratios are sampled from a Gaussian distribution with mean values 2.9 and 4.4, respectively, and standard deviations of .1 times the mean. (The absolute axis lengths are then adjusted to match the volume chosen in (i).)

We analyzed the radial preferences for two different nucleus shapes — spherical and flat ellipsoidal with axis ratios of 1:2:4 — and for the small, medium, and large nuclei with sizes described above.

For each of these six different scenarios we ran 100-200 trials, each with data perturbed as described above and each from a different random starting point. We applied a screening step to remove those trials that have a final objective value greater than

$$f^* + \max(0.5, \min(0.2 \times f^*, 2.0)),$$

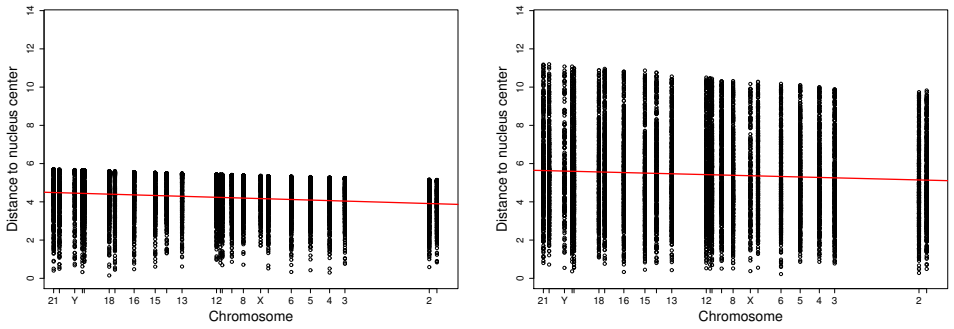
where f^* is the lowest objective value obtained over all trials for this scenario. Table 5.2 shows statistics on the final objective value for each of the six scenarios. Only a few trials were removed in the screening step, mostly for the large spherical nucleus in which the no-overlap solution was not quite attained. After screening, the final objective values were similar for all trials on a given scenario.

Recall that we use Algorithm 2 to solve the convex relaxation (4.9) of the original formulation (4.5), in which the prescribed half-axis lengths (r_{i1}, r_{i2}, r_{i3}) are replaced by the constraints (4.6). We found that a number of the ellipsoids were "more rounded" at the solution than our prescription would require, but that the deformation typically affected only a subset of the ellipsoids and was not too severe. By taking relative difference in the ℓ_2 norm between the vector of actual half-axis lengths and the prescribed values, we found that on small nuclei, an average of 8 of the 46 CTs experienced a relative change of greater than 10%. For medium spherical nuclei, about 7 out of 46 changed by more than 10% while the corresponding number for

TABLE 5.2

Statistics for final objective values attained in the six scenarios, showing number of trials, means, and standard deviations, both before and after the screening step.

		Before Screening			After Screening		
shape	vol (μm^3)	trials	mean	sd	trials	mean	sd
spherical	500	100	3.0889	0.0533	100	3.0889	0.0533
ellipsoidal	500	100	3.2769	0.0660	100	3.2769	0.0660
spherical	1000	200	1.8927	0.4409	195	1.8233	0.0657
ellipsoidal	1000	200	1.9723	0.0714	200	1.9723	0.0714
spherical	1600	100	0.6342	1.4325	89	0.1349	0.0362
ellipsoidal	1600	100	0.1338	0.0291	100	0.1338	0.0291



(a) Spherical Nucluei (b) Ellipsoidal Nuclei

FIG. 5.1. *Scatter plots and regression lines for distances of ellipsoidal chromosome territories to nucleus center, for medium-sized nuclei. Horizontal axis is CT volume, vertical axis is distance to nucleus center.*

large spherical nuclei is 11 out of 46. The statistics for ellipsoidal nuclei are slightly smaller, about 5 for small and large nuclei and 3 for medium nuclei.

We analyzed the solutions generated in the trials remaining after the screening step to find the distances of each ellipsoid from the center of the nucleus. Figure 5.1 contains scatter plots that show the mean volume of each CT (on the horizontal axis) versus the distance between the center of that CT and the nuclear center (on the vertical axis), for a medium-sized nucleus (volume of 1000 μm^3) and for both spherical and flat ellipsoidal shapes. (The scatter plots for the large and small volume nuclei are similar, so we do not show them here.) A least-squares regression line is also shown. In both graphs, a negative trend is detectable, meaning that the larger ellipsoids tend to lie closer to the nuclear center, while the smaller ones prefer peripheral positions. In Table 5.3 we report the slopes of the regression line for all six scenarios, which are negative for each scenario. This is the opposite trend to the one observed in experiments, suggesting that the minimum-overlap criterion alone is insufficient to explain the experimentally observed radial preferences.

TABLE 5.3
Slope of regression line for all six scenarios.

	small	medium	large
spherical	-0.0050562	-0.0029405	-0.0020672
ellipsoidal	-0.0047345	-0.0025045	-0.0018849

5.4. Radial preferences assuming heterologous CT groupings. Khalil et al. [11] showed that CTs tend to assemble in heterologous neighborhoods, causing the distances between homologous chromosome pairs to be larger in general than heterologous inter-CT distances. They discuss a number of possible explanations for this phenomenon, such as that heterologous neighborhoods act as a buffer zone in preventing inter-homologue recombination and protect against the loss of heterozygosity. The authors also analyze whether the radial preferences discussed in the previous subsection could explain the preference for arrangements with larger homologous inter-CT distances. Using simulations, they give a negative answer to this question.

In the following analysis, we invert the question, asking instead whether the

preference for heterologous neighborhoods can explain the observed radial preferences. To investigate this question, we add penalties to our model to discourage the CTs in a homologous pair from being too close to each other. We solve the resulting formulation using a combination of Algorithm 1 for sphere packing with Algorithm 2 for ellipsoid packing.

We denote the set of index pairs (i, j) corresponding to homologous chromosome pairs by H and we introduce a new variable η to capture the proximity of CTs in a homologous pair. Specifically, we define for each ellipsoid i an enclosing sphere that is concentric with the ellipsoid i, with radius λ times the maximum semi-axis length r_{i1} of the CT, where $\lambda \geq 1$ is a user-defined parameter. We define η to be the maximal overlap of these enclosing spheres, over all homologous pairs, by adding constraints whose form is similar to (3.4b). We then add a penalty term $c\eta$ to the objective (for some penalty parameter $c \geq 0$) to obtain the following extension of formulation (4.5).

$$\min_{\xi,\eta,(c_i,S_i,\Sigma_i),i=1,2,\ldots,N} \quad \xi + c\eta \tag{5.1a}$$

$$\text{subject to} \quad \xi \geq \hat{O}(c_i, c_j, \Sigma_i, \Sigma_j), \qquad 1 \leq i < j \leq N, \tag{5.1b}$$

$$\lambda(r_{i1} - r_{j1}) - \|c_i - c_j\|_2 \leq \eta, \qquad (i,j) \in H, \tag{5.1c}$$

$$\mathcal{E}_i \subset \mathcal{E}, \qquad i = 1, 2, \ldots, N, \tag{5.1d}$$

$$\Sigma_i = S_i^2, \tag{5.1e}$$

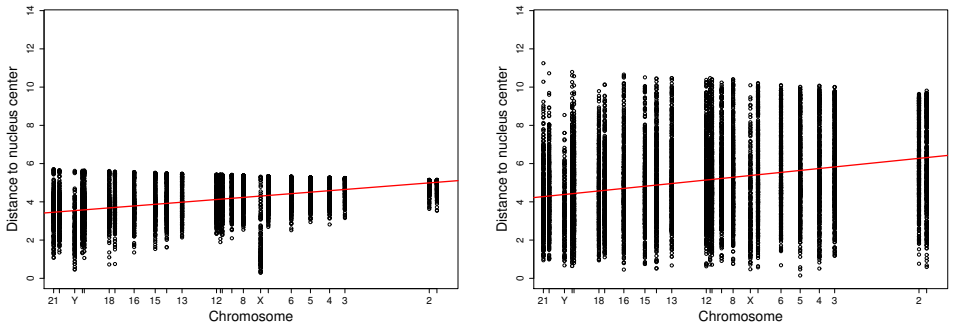
$$\text{semi-axes of } \mathcal{E}_i \text{ are } r_{i1}, r_{i2}, r_{i3}, \qquad i = 1, 2, \ldots, N. \tag{5.1f}$$

We relax this formulation in the style of (4.9). To solve, we extend Algorithm 2 by adding linearizations of the constraints (5.1b) to each subproblem, as in (3.6b).

For our simulations, we choose $c = 100$ and $\lambda = 1.25$. As in Subsection 5.3, we generated about 100-200 trials by perturbing CT volumes and dimensions randomly around given mean values and by using different random starting points. The screening procedure described in the previous subsection was applied to remove those trials with less competitive final objective values.

Statistics for the final objectives are shown in Table 5.4. The large objective values in the first line of the table indicates that for small spherical nuclei, it was not possible to find solutions in which the homolog separation was enforced adequately. (The only trial that survived screening was one that violated these conditions significantly less than most others.) Among the other scenarios, only the medium spherical nucleus saw significant numbers of trials removed by screening. Here, most of the trials attained final objectives quite close to 1.90, while the others had significantly higher values. In the other four scenarios — small ellipsoidal, medium ellipsoidal, large spherical, and large ellipsoidal — proximity penalties for homologous pairs were not incurred, and final objective values were tightly clustered.

The convex relaxation of our problem that encourages separation of homologous CT pairs does less well in preserving the dimensions of the ellipsoids than the formulation considered in Section 5.3. For spherical nuclei 26 out of the 46 CTs for small nuclei experienced a relative change in the half-axes lengths of more than 10%. For the medium spherical nuclei it was in average 11 out of 46 and for the large spherical nuclei 17 out of 46. The statistics for the ellipsoidal nuclei were somewhat smaller: 12 for the small nuclei, 4 for the medium nuclei, and 9 for the large nuclei. On the small nuclei, the distortions can be explained by the tightness of space, while on large nuclei, the fact that all CTs can be fit without any overlap reduces the need for them to



(a) Spherical Nuclei (b) Ellipsoidal Nuclei

FIG. 5.2. *Scatter plots and regression lines for ellipsoidal chromosome territory distances to nucleus center, where penalties to enforce heterologous groupings are present in the formulation. Horizontal axis is CT volume, vertical axis is distance to nucleus center.*

adopt their lowest-volume dimensions (which would achieve the prescribed semi-axis lengths).

Figure 5.2 contains scatter plots showing the mean volume of each CT (on the horizontal axis) plotted against the distance between the center of that CT and the nuclear center (on the vertical axis), for a medium-sized nucleus (volume of 1000 μm^3) and for both spherical and flat ellipsoidal shapes. (As in Subsection 5.3, scatter plots for the large and small volume nuclei are similar, so we do not show them here.) Here the regression line shows a significant positive trend, meaning that the smaller ellipsoids tend to lie in the interior of the nucleus, while the larger ones prefer peripheral positions. Hence, by adding penalties for nearby homologous pairs to our formulation, we obtain a considerable improvement in our ability to match the radial preferences observed in nature.

Another interesting observation, more evident in Figure 5.2(a), is that the X and Y chromosomes both lie closer to the nucleus center than their size would suggest. This makes sense, as these are the only two chromosomes not subject to the homologous-pair separation penalties.

In Table 5.5 we report the slopes of the regression line for all six size / shape scenarios considered in this section. These results highlight a significant difference between spherical and flat-ellipsoidal nuclei: The radial preference is consistently

TABLE 5.4

Statistics for final objective values attained in the six scenarios in which homolog proximity is penalized, showing number of trials, means, and standard deviations, both before and after the screening step.

		Before Screening			After Screening		
shape	vol (μm^3)	trials	mean	sd	trials	mean	sd
spherical	500	100	294.0617	46.8185	1	184.0292	0.0000
ellipsoidal	500	100	3.6556	0.2691	95	3.5987	0.0879
spherical	1000	200	15.0885	20.7260	114	1.8993	0.0833
ellipsoidal	1000	200	2.0088	0.5118	192	1.9060	0.0691
spherical	1600	100	0.4424	1.2293	94	0.1369	0.0213
ellipsoidal	1600	100	0.2752	0.8097	97	0.1343	0.0251

weaker for spherical nuclei than for flat-ellipsoidal nuclei.

TABLE 5.5
Slope of regression line for all six scenarios assuming heterologous CT groupings.

	small	medium	large
spherical	0.0031803	0.0078299	0.0072028
ellipsoidal	0.0081768	0.0102088	0.0145001

6. Discussion. We have described a bilevel optimization procedure for finding local solutions of the problem of packing spheres and ellipsoids in finite volumes with minimal overlap. We used these procedures to model and analyze chromosome arrangement in cell nuclei. Semidefinite programming duality is used to obtain the sensitivity information needed to construct the approximation to the upper-level problem that is solved at each iteration of the trust-region procedure. Our convergence analysis takes place in a general setting in which the lower-level problems are semidefinite programs parametrized by their objective coefficient matrix; it is not confined to the specific form of the semidefinite programs arising from the S-procedure for overlapping ellipsoids. Thus, our trust-region framework may be adaptable to other design problems involving parametrized systems that can be modeled by semidefinite programs.

In the CT packing application discussed in Section 5, we initially found that the arrangements observed experimentally could not be explained by the simple geometric principle of minimizing the maximum overlap. However, when we enhanced the model to capture the recently observed phenomenon of heterologous neighborhoods / homologous pair separation, the radial preferences observed in nature (in which larger CTs tended to lie further from the nuclear center) were recovered in our simulations. The homologous-pair-separation aspects of our model are governed by two positive parameters c and λ; we reported results in Subsection 5.4 only for the values $c = 100$ and $\lambda = 1.25$. From an examination of Tables 5.3 and 5.5, we speculate that it would be possible to choose these parameters in such a way that the slope of the regression line for spherical nuclei would be approximately zero, while the corresponding slope for ellipsoidal nuclei would be positive. Such a result would be consistent with experimental observations that identify no clear radial preference for spherical nuclei, but a pronounced radial preference for ellipsoidal nuclei.

We obtained results on a limited but representative range of nucleus dimensions. In future work, we will explore CT configurations for a wider range of ellipsoidal shapes and sizes, corresponding to known dimensions of nuclei in different cell types. We will also enhance the model as further biological results are obtained, aiming to find biologically plausible, elementary principles that explain experimental observations (in the spirit of Occam's Razor).

Appendix A. Parametrized Semidefinite Programs: Technical Results. In this section we consider the following primal-dual pair of semidefinite programs that are parametrized by the primal objective term C:

$$t(C) := \min_X \langle C, X\rangle \;\text{ s.t. }\; \langle A_i, X\rangle = b_i, \;\; i = 1, 2, \ldots, p, \;\; X \succeq 0, \tag{A.1}$$

$$\max_{\zeta, S} b^T\zeta \;\text{ s.t. }\; \sum_{i=1}^{p} \zeta_i A_i + S = C, \;\; S \succeq 0. \tag{A.2}$$

We denote solutions of these problems by $X(C)$ and $(\zeta(C), S(C))$, respectively. (Our interest is in the application to the SDP pair (4.11), (4.15), but we have simplified the notation here.)

We show first that the solutions to (A.1) are uniformly bounded in a neighborhood of a C for which a strictly feasible point for the dual (A.2) exists. The result is an almost immediate consequence of [23, Theorem 4.1].

LEMMA A.1. *Consider the primal-dual pair (A.1), (A.2) of semidefinite programs: Suppose that (A.1) is feasible (with feasible point $\hat{X}$), and that at some matrix $C_0 \in S\mathbb{R}^{n\times n}$, there exists a strictly feasible point $(\hat{\zeta}, \hat{S})$ for (A.2), where the eigenvalues of $\hat{S}$ are bounded below by $\sigma > 0$. Then there exists a constant $\delta > 0$ such that for all matrices $C \in S\mathbb{R}^{n\times n}$ with $\|C - C_0\|_F \le \delta$, (A.1) has a nonempty solution set, and all solutions $X(C)$ are bounded as follows:*

$$\|X(C)\|_F \le \frac{2}{\sigma}(\langle \hat{X}, \hat{S}\rangle + \delta\|\hat{X}\|_F).$$

Moreover the optimal values of the problems (A.1) and (A.2) are equal.

Proof. We have for any X feasible for the primal (note that the primal feasible region does not depend on C) that

$$\begin{aligned}\langle C, X\rangle &= \langle C_0, X\rangle + \langle C - C_0, X\rangle \\ &= \langle \sum_{i=1}^{p} \hat{\zeta}_i A_i + \hat{S}, X\rangle + \langle C - C_0, X\rangle \\ &= b^T\hat{\zeta} + \langle \hat{S}, X\rangle + \langle C - C_0, X\rangle.\end{aligned}$$

Since $b^T\hat{\zeta}$ is independent of X, we can obtain an equivalent to the primal problem by replacing its objective $\langle C, X\rangle$ by $\langle \hat{S} + C - C_0, X\rangle$. Using the assumed feasible point $\hat{X}$ of (A.1) (note that there is no dependence of $\hat{X}$ on C), we can formulate (A.1) equivalently as follows:

$$\min_X \langle \hat{S} + C - C_0, X\rangle \ \text{ s.t. } \ \langle A_i, X\rangle = b_i, \ \ i = 1, 2, \ldots, p, \ \ X \succeq 0, \tag{A.3a}$$

$$\langle \hat{S} + C - C_0, X\rangle \le \langle \hat{S} + C - C_0, \hat{X}\rangle. \tag{A.3b}$$

By choosing $\delta \in (0, \sigma/2]$, we have that all eigenvalues of $\hat{S} + C - C_0$ are bounded below by $\sigma/2$. Hence from "Fact 14" of [23], we have that

$$\langle \hat{S}+C-C_0, X\rangle \le \langle \hat{S}+C-C_0, \hat{X}\rangle \ \Rightarrow \ \|X\|_F \le \frac{2}{\sigma}\langle \hat{S}+C-C_0, \hat{X}\rangle \le \frac{2}{\sigma}(\langle \hat{S}, \hat{X}\rangle + \delta\|\hat{X}\|_F).$$

Hence, (A.3) involves the minimization of a continuous function over a nonempty compact set, so the solution set exists, and all solutions are bounded as claimed. The last claim can be derived exactly as in [23, Theorem 4.1]. □

The next result examines the solution of a sequence of parametrized SDPs.

THEOREM A.2. *Given A_i, $i = 1 = 1, 2, \ldots, p$ and $b \in \mathbb{R}^p$ as in Lemma A.1, such that (A.1) is feasible, let $\bar{C} \in S\mathbb{R}^{n\times n}$ be such that there exists a strictly feasible point for (A.2) when $C = \bar{C}$. Consider a sequence $\{C_k\}$ with $C_k \in S\mathbb{R}^{n\times n}$ and $C_k \to \bar{C}$. Then the following is true:*

(i) There is a constant $\beta > 0$ and index K such that (A.1) with $C = C_k$ has nonempty solution set for all $k \ge K$, and $\|X(C_k)\|_F \le \beta$ for all such solutions.

(ii) If $\{X(C_k)\}$ is a sequence of solutions of (A.1) with $C = C_k$, then this sequence has at least one accumulation point, and all such accumulation points are solutions of (A.1) with $C = \bar{C}$.

Proof. The first claim (i) is an immediate consequence of Lemma A.1. For (ii), note that boundedness of $X(C_k)$ ensures existence of accumulation points. Suppose that $\bar{X}$ is such a point and assume WLOG that $X(C_k) \to \bar{X}$. Note first that $\bar{X}$ is feasible for (A.1) regardless of C. If $\bar{X}$ were not optimal for (A.1) with $C = \bar{C}$, then there would exist another feasible matrix $\tilde{X}$ with $\langle \bar{C}, \tilde{X} \rangle < \langle \bar{C}, \bar{X} \rangle$. But since

$$\lim_k \langle C_k, \tilde{X} \rangle = \langle \bar{C}, \tilde{X} \rangle < \langle \bar{C}, \bar{X} \rangle = \lim_k \langle C_k, X(C_k) \rangle,$$

we have that $\langle C_k, \tilde{X} \rangle < \langle C_k, X(C_k) \rangle$ for all k sufficiently large, contradicting optimality of $X(C_k)$. Hence (ii) is true. □

We next prove some elementary and useful facts about the value function $t(C)$ of (A.1).

LEMMA A.3. *Suppose that (A.1) is feasible, and let $\bar{C} \in S\mathbb{R}^{n\times n}$ be such that there exists a strictly feasible point for (A.2) when $C = \bar{C}$. Then there exists a neighborhood $\mathcal{N}$ of $\bar{C}$ within which the following claims are true.*

(i) $t(\cdot)$ is a concave function.

(ii) For all $C \in \mathcal{N}$ and all $\Delta C \in S\mathbb{R}^{n\times n}$, we have

$$t(C + \Delta C) \leq t(C) + \langle X(C), \Delta C \rangle, \tag{A.4}$$

where $X(C)$ is any solution of (A.1).

(iii) Any $X(C)$ that solves (A.1) belongs to the Clarke subdifferential of $t(\cdot)$ at C.

(iv) $t(\cdot)$ is Lipschitz continuous in $\mathcal{N}$.

Proof. The proof of (i) is elementary.

For (ii), note first from Lemma A.1 that we can choose $\mathcal{N}$ so as to ensure that a solution $X(C)$ to (A.1) exists for all $C \in \mathcal{N}$. We have (denoting by Ψ the feasible set for (A.1)) that

$$t(C + \Delta C) = \min_{X \in \Psi} \langle C + \Delta C, X \rangle \leq \langle C + \Delta C, X(C) \rangle = t(C) + \langle \Delta C, X(C) \rangle,$$

as required.

For (iii), note that (by taking the negative of the inequality (A.4)) $-X(C)$ is a subgradient of the convex function $(-t)(C)$, and so $-X(C)$ belongs to the Clarke subdifferential of $(-t)(C)$. It follows from [2, p. 128, Exercise 8(c)] (with $\lambda = -1$) that $X(C)$ belongs to the Clarke subdifferential of $t(C)$, as claimed.

For (iv), note from Lemma A.1 that we can choose $\mathcal{N}$ such that $\|X(C)\|$ is uniformly bounded for all $C \in \mathcal{N}$ (by $\beta > 0$, say). Denoting by C_1 and C_2 any two points in $\mathcal{N}$, we have from (A.4) that

$$\begin{aligned} t(C_2) &\leq t(C_1) + \langle X(C_1), C_2 - C_1 \rangle, \\ t(C_1) &\leq t(C_2) + \langle X(C_2), C_1 - C_2 \rangle. \end{aligned}$$

Thus

$$|t(C_1) - t(C_2)| \leq \max\left(\|X(C_1)\|, \|X(C_2)\|\right) \|C_1 - C_2\| \leq \beta \|C_1 - C_2\|,$$

proving the Lipschitz property. □

Appendix B. Strong Duality for the Overlap Problem. Here we give a brief justification of the strong duality relationship between (4.2) and (4.4). For the purpose of this argument, we identify (4.2) with the standard-form dual problem (A.2) and (4.4) with the standard-form primal problem (A.1). Note that we have a strictly feasible point for (4.4) (thus (A.1)), as discussed at the end of Subsection 4.1. We denote by p^* the optimal value of (A.1), with $p^* = -\infty$ when the problem is infeasible, and we use d^* to denote the optimal value of (A.2), with $d^* = \infty$ when this problem is infeasible. The following result is proved by de Klerk [7, Theorem 2.2].

THEOREM B.1. *Assume that $p^* > -\infty$ and that (A.1) is strictly feasible. Then $p^* = d^*$ and the solution to (A.2) is attained.*

The following corollary is easily proved.

COROLLARY B.2. *Suppose that (A.1) is strictly feasible. Then $p^* = d^*$, and the solution to (A.2) is attained when $p^* > -\infty$.*

Proof. Consider first the case of $d^* > -\infty$. Then (A.2) is feasible, and by weak duality we have $p^* \geq d^*$. Since $p^* > -\infty$, we can apply Theorem B.1 to deduce that $p^* = d^*$ in this case, and the solution to (A.2) is attained.

Now consider $d^* = -\infty$, in which case (A.2) is infeasible. If we were to have $p^* > -\infty$, then Theorem A.2 would imply that (A.2) attains a solution, which is not the case since this problem is not even feasible. Hence $p* = -\infty$. □

Appendix C. Proofs of Technical Lemmas from Subsection 4.3.

Proof. (Lemma 4.1) We appeal to results about the Clarke subdifferential applied to max-functions and sums of functions. First, note that the strict interiority assumption means that we can apply Lemma A.3 (iv) to deduce that each t_l^* is Lipschitz continuous in a neighborhood of $\bar{C}$. Hence, applying [2, Theorem 6.1.5], we have that

$$\partial\, t_{\mathcal{F}}^*(C) \subset \mathrm{conv}\{\partial t_l^*(\bar{C}) \,:\, l \in \bar{\mathcal{F}}(\bar{C})\}, \tag{C.1}$$

where $\mathrm{conv}(\cdot)$ denotes the convex hull. The unnumbered corollary on p. 52 of [5] can be used to show that when $\bar{C}$ is a solution of (4.19), we have

$$0 \in \partial\, t_{\mathcal{F}}^*(\bar{C}) + N_\Omega(\bar{C}).$$

The result follows by combining this expression with (C.1). □

Proof. (Lemma 4.2) (i) If $r(\mathcal{F}, \bar{C}, M_{\mathcal{F}}(\bar{C}), \rho) = t_{\mathcal{F}}^*(\bar{C})$ for some $\rho > 0$, then $\Delta C = 0$ achieves the minimum in (4.22), for $C = \bar{C}$. Hence, there exist μ_l, $l \in \mathcal{F}$ such that the optimality conditions (4.23) are satisfied with $\Delta C = 0$ and $\tau = 0$. Thus, conditions (4.21) are satisfied with $\bar{M}_l = \bar{M}_l(\bar{C})$ and the same values of μ_l, $l \in \mathcal{F}$.

(ii) Trivial, as the size of the feasible region for $L(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho)$ increases with ρ.

(iii) We need to show that for ρ_1 and ρ_2 with $0 < \rho_1 < \rho_2$, we have

$$\frac{\Lambda(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho_1)}{\rho_1} \geq \frac{\Lambda(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho_2)}{\rho_2}.$$

Using the formulation (4.25) of $L(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho)$, and in particular the convex function $G_{\mathcal{F}}(\cdot; C, M_{\mathcal{F}}(C))$ defined in (4.24), we have that

$$G_{\mathcal{F}}(0; C, M_{\mathcal{F}}(C)) = \max_{l \in \mathcal{F}} \langle C, M_l(C) \rangle = t_{\mathcal{F}}^*(C).$$

Given the solution $\Delta C(\rho_2)$ of $L(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho_2)$, note that $\frac{\rho_1}{\rho_2}\Delta C(\rho_2)$ is feasible for $L(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho_1)$. Since $\Delta C(\rho_1)$ is optimal for $L(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho_1)$, we have

$$\begin{aligned} G_{\mathcal{F}}(\Delta C(\rho_1); C, M_{\mathcal{F}}(C)) &\leq G_{\mathcal{F}}\left(\frac{\rho_1}{\rho_2}\Delta C(\rho_2); C, M_{\mathcal{F}}(C)\right) \\ &\leq \left(1 - \frac{\rho_1}{\rho_2}\right) G_{\mathcal{F}}(0; C, M_{\mathcal{F}}(C)) + \frac{\rho_1}{\rho_2} G_{\mathcal{F}}(\Delta C(\rho_2); C, M_{\mathcal{F}}(C)). \end{aligned}$$

The result follows by rearrangement of this expression, since

$$\begin{aligned} \Lambda(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho_1) &= G_{\mathcal{F}}(0; C, M_{\mathcal{F}}(C)) - G_{\mathcal{F}}(\Delta C(\rho_1); C, M_{\mathcal{F}}(C)), \\ \Lambda(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho_2) &= G_{\mathcal{F}}(0; C, M_{\mathcal{F}}(C)) - G_{\mathcal{F}}(\Delta C(\rho_2); C, M_{\mathcal{F}}(C)). \end{aligned}$$

(iv) The result follows immediately from Lemma A.3 (iv), when we use the definition (4.18) and the fact that

$$r(\mathcal{F}, C, M_{\mathcal{F}}(C), \rho) = \max_{l \in \mathcal{F}} t_l^*(C) + \langle \Delta C(\rho), M_l(C) \rangle.$$

□

Proof. (Lemma 4.4) Note first that since Ω is compact, the set $\Psi_l(\bar{t}) := \{C \in \Omega \mid t_l^*(C) \geq \bar{t}\}$ is also compact, for any $l \in \{1, 2, \ldots, m\}$ and any $\bar{t} > 0$. Since $t_l^*(\cdot)$ is continuous at every point of this set, under the stated assumptions, it is uniformly continuous on this set. Thus for any $\epsilon > 0$, there is a value $\delta = \delta_l(\epsilon) > 0$ such that (i) holds. By defining $\delta := \min_{l=1,2,\ldots,m} \delta_l(\epsilon)$, we obtain (i).

For (ii), we suppose for contradiction that for some $\epsilon > 0$, there is no $\delta > 0$ with the property claimed. Thus, for any sequence $\{\delta_r\}$ with $\delta_r \downarrow 0$, we can find $\bar{C}_r \in \Omega$, $C_r \in \Omega$ with $\|C_r - \bar{C}_r\| \leq \delta_r$, and $l \in \{1, 2, \ldots, m\}$ such that

$$t_l^*(C_r) > \max(0, t_l^*(\bar{C}_r)) + \epsilon. \tag{C.2}$$

By taking a subsequence if necessary, we can assume that this inequality holds for some fixed $l \in \{1, 2, \ldots, m\}$. Since all C_r belong to the compact set $\Omega \cap \{C \mid t_l^*(C) \geq \epsilon\}$, we can assume (by taking another subsequence if necessary) that $C_r \to \hat{C}$, for some $\hat{C}$ with $t_l^*(\hat{C}) \geq \epsilon$. It follows that $\bar{C}_r \to \hat{C}$ also, so by using continuity of t_l^* and taking limits in both sides of (C.2), we obtain

$$t_l^*(\hat{C}) = \lim_{r \to \infty} t_l^*(C_r) \geq \lim_{r \to \infty} \max(0, t_l^*(\bar{C}_r)) + \epsilon = \max(0, t_l^*(\hat{C})) + \epsilon \geq t_l^*(\hat{C}) + \epsilon,$$

a contradiction. □

Acknowledgements.

We thank Saira Mian for helpful discussions about the application to chromosome arrangement in cell nuclei, and Henry Wolkowicz for providing the strong duality argument of Appendix B. We are also grateful to two referees of the original version, who made highly perceptive and constructive comments. Finally, we thank the Institute for Mathematics and its Applications at the University of Minnesota for supporting visits by both authors while this research was conducted.

REFERENCES

[1] A. Bolzer, G. Kreth, I. Solovei, D. Koehler, K. Saracoglu, C. Fauth, S. Müller, R. Eils, C. Cremer, M. R. Speicher, and T. Cremer. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.*, 3, 2005.

[2] J. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples*. CMS Books in Mathematics. Springer, 2000.

[3] S. Boyd, L. El Ghoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in Systems and Control Theory*. SIAM Publications, Philadelphia, Penn., 1994.

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2003.

[5] F. H. Clarke. *Optimization and Nonsmooth Analysis*. John Wiley, New York, 1983.

[6] T. Cremer and M. Cremer. Chromosome territories. *Cold Spring Harb. Perspect. Biol.*, 2010.

[7] E. de Klerk. *Aspects of Semidefinite Programming: Interior-Point Algorithms and Selected Applications*, volume 76 of *Applied Optimization*. Kluwer, 2002.

[8] A. Donev, I. Cisse, D. Sachs, E. A. Variano, F. H. Stillinger, R. Connelly, S. Torquato, and P. M. Chaikin. Improving the density of jammed disordered packings using ellipsoids. *Science*, 303:990–993, February 2004.

[9] M. Grant and S. Boyd. *CVX User's Guide*. Stanford University, version 1.22 edition, February 2012.

[10] T. A. Hales. A proof of the Kepler conjecture. *Annals of Mathematics, Second Series*, 162(3):1065–1185, November 2005.

[11] C. Heride, M. Ricoul, K. Kiêu, J. von Hase, V. Guillemot, C. Cremer, K. Dubrana, and L. Sabatier. Distance between homologous chromosomes results from chromosome positioning constraints. *J. Cell Sci.*, 123:4063–4075, 2010.

[12] A. Khalil, J. L. Grant, L. B. Caddle, E. Atzema, K. D. Mills, and A. Arneodo. Chromosome territories have a highly nonspherical morphology and nonrandom positioning. *Chromosome Res.*, 15:899–916, 2007.

[13] B. D. Lubachevsky and R. L. Graham. Curved hexagonal packings of equal disks in a circle. *Discrete and Computational Geometry*, 18(2):179–194, June 2007.

[14] N. V. Marella, S. Bhattacharya, L. Mukherjee, J. Xu, and R. Berezney. Cell type specific chromosome territory organization in the interphase nucleus of normal and cancer cells. *J. Cell. Physiol.*, 221:130–138, 2009.

[15] N. V. Marella, B. Seifert, P. Nagarajan, S. Sinha, and R. Berezney. Chromosomal rearrangements during human epidermal keratinocyte differentiation. *J. Cell. Physiol.*, 221:139–146, 2009.

[16] I. Müller, S. Boyle, R. H. Singer, W. A. Bickmore, and J. R. Chubb. Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells. *PLoS One*, 5, 2010.

[17] Y. Nesterov and A. S. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming*. SIAM Publications, Philadelphia, 1994.

[18] I. Pólik and T. Terlaky. A survey of the S-Lemma. *SIAM Review*, 49(3):371–418, 2007.

[19] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, N.J., 1970.

[20] C. A. Rogers. *Packing and Covering*, volume 54 of *Cambridge Tracts in Mathematics and Mathematical Physics*. Cambridge University Press, 1964.

[21] H. Tanabe, S. Müller, M. Neusser, J. von Hase, E. Calcagno, M. Cremer, I. Solovei, C. Cremer, and T. Cremer. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Natl. Acad. Sci.*, 99:4424–4429, 2002.

[22] A. Thue. Über die dichteste Zusammenstellung von kongruenten Kreisen in einer Ebene. *Norske Vod. Selsk. Skr.*, 1:1–9, 1910.

[23] M. J. Todd. Semidefinite optimization. *Acta Numerica*, 10:515–560, 2001.

[24] M. J. Zeitz, L. Mukherjee, S. Bhattacharya, J. Xu, and R. Berezney. A probabilistic model for the arrangement of a subset of human chromosome territories in WI38 human fibroblasts. *J. Cell. Physiol.*, 221:120–129, 2009.